The University of Southern Mississippi

The Aquila Digital Community

Dissertations

Spring 5-1-2018

Personality in California Sea Lions (*Zalophus californianus*) and Harbor seals (*Phoca vitulina*): Methodological Convergence and Species-Specific Emotional Repertoires

Amber de Vere
University of Southern Mississippi

Personality in California Sea Lions (*Zalophus californianus*) and Harbor seals (*Phoca vitulina*): Methodological Convergence and Species-Specific Emotional Repertoires.

by

Amber J. de Vere

A Dissertation
Submitted to the Graduate School,
the College of Education and Psychology
and the Department/ School of Psychology
at The University of Southern Mississippi
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

Approved by:

Dr. Lucas Keefer, Committee Chair
Dr. Lauren Highfill
Dr. Richard Mohn
Dr. Donald Sacco

Dr. Lucas Keefer Committee Chair	Dr. Joe Olmi Department Chair	Dr. Karen S. Coats Dean of the Graduate School

May 2018

Published by the Graduate School

THE UNIVERSITY OF SOUTHERN MISSISSIPPI

ABSTRACT

Despite the wide acceptance of animal personality as a valid area of study, research on marine mammal personality remains remarkably scarce. What literature does exist predominantly focuses on bottlenose dolphins (Frick, 2016; Highfill & Kuczaj, 2007; Kuczaj, Highfill & Byerly, 2012; Lilley, de Vere, Yeatre & Kuczaj, 2018; Moreno, Highfill & Kuczaj, 2017). There is also strong evidence for individual differences in grey seals (Robinson et al., 2015; Twiss & Franklin, 2010; Twiss, Culloch & Pomeroy, 2011; Twiss, Cairns, Culloch, Richards & Pomeroy, 2012), and preliminary research has provided evidence of broad personality factors in pinniped species using behavioral coding (de Vere, Lilley & Highfill, 2017) and trait rating methods (Ciardelli, Weiss, Powell & Reiss, 2017). Several aspects of personality are not well documented across many taxa, including age-related patterns, the species-relevance of emotional trait words, potential issues associated with the non-human Dominance factor, and the convergent validity of multiple methods. The current study therefore aimed to address these issues in two pinniped species, California sea lions and harbor seals, and provides the first cross-method validation of personality dimensions in these taxa. There was some evidence that pinniped trainers could reliably rate the emotional states experienced by these species. Trait rating assessments produced three personality factors for each species; these exhibited good cross-method convergence in California sea lions, but not harbor seals. Dominance rankings were correlated with one behavioral and one rating factor in each species, although this was somewhat confounded by the extremely close correspondence of dominance and age.

ACKNOWLEDGMENTS

This research would not have been possible without the support of the pinniped trainers at Six Flags Discovery Kingdom, Vallejo, and the animals under their care.

Dr. Lauren Highfill and Dr. Lucas Keefer were instrumental to the completion of this project, as they both provided exceptional support beyond that expected of academic advisors.

.

DEDICATION

This dissertation is dedicated to Dr. Stan Kuczaj. He is the reason this research was made possible, and inspired not only this particular project, but so much of my research. I and his other students will continue to do our best to honor his legacy.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER I - INTRODUCTION

That non-human animals have personality is no longer a controversial statement, although this has not always been the case. Many labels, including personality, temperament, and behavioral syndromes, have been used to describe essentially the same phenomenon: individual differences in behavior that are consistent across time and contexts (Gosling, 2001). Animal personality research has expanded not only through cross-population and -study replication within a single species (e.g. chimpanzees: King & Figueredo, 1997; King, Weiss & Farmer, 2005; Weiss et al., 2007; Weiss, King & Hopkins, 2009), but also through assessments of increasingly diverse taxa (Gosling & John, 1999). This variety has permitted examination of the cross-species generality of underlying dimensions, providing insights into the evolution of personality.

The most recent review to compare the non-human animal literature to the human Five Factor Model of personality found that an Extraversion-like factor occurred most frequently across the 12 studied species (e.g. chimpanzees, dogs, pigs), although the labels and behavioral content of these factors varied across species (Gosling & John, 1999). Neuroticism and Agreeableness were the next most commonly replicated dimensions, followed by Openness. In contrast, a clear Conscientiousness factor was found only in chimpanzees, although bottlenose dolphins have since been reliably rated on this factor (Highfill & Kuczaj, 2007; Moreno, 2017), and a combined factor with Openness was present in cats and dogs (Gosling & John, 1999). There was substantial evidence for Dominance as a separable non-human animal factor, rather than a lower-level facet subsumed within another factor, such as Extraversion (DeYoung, Quilty & Peterson, 2007). Since this review, animal personality research has only continued to expand (e.g. Gosling, 2001; Vazire, Gosling, Dickey & Schapiro, 2007), providing strong evidence for personality in a range of taxa. In addition, cross-species comparisons provide a window into the evolution of personality, as well as unique ways in which traits are expressed in different species.

Methods & Convergent Validity

Two primary methods are used to assess animal personality. The first, behavioral coding, involves making observations of animals and recording behaviors that are often selected from ethograms of species-specific behaviors (Freeman, Gosling & Schapiro, 2011). These observations can be carried out in a naturalistic setting, with no human manipulation, or during experimental testing. Trait rating is the major alternative method, in which the tendencies of individuals are judged on trait words typically selected from a pre-existing model of personality. This method originates from human personality research, in which perceiver ratings have been increasingly validated (Vazire & Carlson, 2010). In non-human research, raters are usually people who have long term experience with the focal animals, such as trainers and pet owners. Alternatively, they may be inexperienced with the subjects, and make ratings after observing the animals in a certain setting, such as during a veterinary examination or behavioral tests (Freeman et al., 2011).

In order to assess the reliability of data obtained from trait rating, more than one rater will judge each animal. As a result, the identity of the raters must be taken into account. For example, ratings have previously been found to be unreliable where raters had different experiences with the subjects (Highfill, Hanbury, Kristiansen, Kuczaj & Watson, 2010). As well as the contexts in which raters have experience with an animal, it is also important to consider the length of their experience. In many trait rating studies, a criteria is used to determine whether raters have a sufficient amount of experience (e.g. Horback, Miller & Kuczaj, 2013; Lloyd, Martin, Bornett-Gauci & Wilkinson, 2007; Lloyd, Martin, Bornett-Gauci & Wilkinson, 2008). For example, such criteria may require six months of repeated daily encounters, one year of experience, or living with/providing care for an animal for at least two years (Horback et al., 2013; Lloyd et al., 2008; Morris, Doe & Godsell, 2008). In other studies, raters may have made just a few weeks of concentrated observations (e.g. Capitanio, 1999). A few assessments have compared the reliability of ratings made by humans who are familiar with subject animals with those who are unfamiliar. For example, students who had observed a group of Japanese macaques for at least two hours a day for a month produced reliable ratings of these animals, while students who had observed the group for less than five hours did not (Martau, Caine & Candland, 1985). More recently, the inter-rater reliability of ratings made by students who observed macaques for five consecutive days and those made by eight ‘experts’ who had worked with the macaque group for between several months and a few years were almost identical (Uher, Werner & Gosselt, 2013). Reliable ratings have also been produced by previously inexperienced raters after making three months of concentrated observations (Feaver, Mendl & Bateson, 1986).

Therefore, there is great variation both within and between studies, and no clear consensus on the amount of experience necessary for someone to be able to accurately or reliably rate an animal’s personality. It seems to be assumed that the longer the experience, the better, but arguments can be made for both sides. For instance, longer experience is likely to involve greater exposure to an animal, which may facilitate more accurate, representative ratings. However, extended experience may also be a source of potential bias, due to familiarity and the potential for increased inter-dependence of raters. Furthermore, incorporating measures beyond inter-rater reliability could provide a more detailed picture of the ways in which the extent of human-animal experience affect trait ratings. Examining whether raters who have spent less time in contact with individual animals produce less reliable ratings, or have greater uncertainty in their ratings, would therefore be a useful addition to rating studies in general. This could be achieved by comparing raters with differing lengths of experience on their ratings of the same animal, or by comparing the ratings made of younger animals, with whom raters will have inherently less experience, with those of adult animals.

Finally, some argue that using a combination of behavioral coding and trait rating is most informative for determining the accuracy of assessments of animal personality (Vazire, Gosling, Dickey & Schapiro, 2007). The reliability of measures can be assessed using the agreement between observers or raters. However, single-method assessments rarely estimate the extent to which variation stems from method-specific measurement error versus individual differences on the target personality construct. Using more than one method to determine the extent to which results from each method predict or are associated with each other can be useful for partitioning out these potential sources of variance (Freeman et al., 2011). In human research, multitrait-multimethod matrices have been incorporated into such studies to account for the variance associated with methodology (e.g. Watson, Suls & Haig, 2002), but the sample requirements for such analyses are not often achievable in non-human personality research. As a result, variance specifically associated with the methods used does not tend to be estimated, with construct-related variance forming the focus of these studies. For example, trait rating assessments of chimpanzee personality have shown strong correspondence with relevant behaviors, such as low scores on the dimension of Dominance with behaviors like touching and playing with other animals (Pederson, King & Landau, 2005). A combination of methods can therefore be useful for providing information about the extent to which variation is attributable to method variance versus the target constructs.

Marine Mammal Personality

Despite the ever-increasing range of taxa appearing in the personality literature, very few marine mammal species to date have been assessed for personality. By far the best studied of these species is the bottlenose dolphin. In the first study of this kind, trainers rated bottlenose dolphins in a captive population on two occasions, more than one year apart (Highfill & Kuczaj, 2007). Ratings for each of the human Five Factors were reliable, and individual dolphins showed reasonable stability across the inter-test interval, despite being displaced by hurricane Katrina during this time. Bottlenose dolphins have also been assessed on a subset of traits, across several interactive contexts (Highfill, Kuczaj & Byerly, 2012). Again, ratings were reliable, although there were individually specific patterns of behavioral consistency across contexts. More recent evidence further supports the utility of the trait rating method in this species, with another dolphin population rated reliably on Extraversion, Neuroticism, Openness, and Conscientiousness factors (Moreno, 2017). In the same population, behavioral coding has also been used to cluster behaviors into broader personality traits (Frick, 2016). Some composite traits emerged across all dolphins, but their content differed between sexes. For example, curiosity contained bubble bursts, caused by rapid blowhole exhalation, in both sexes, but male curiosity also contained approaches to other dolphins, while female curiosity involved orienting towards objects. Some traits were also unique to each sex, such as Sexual in males and Affiliative in females. The first evidence for convergent validity of personality across methods in any marine mammal has also recently been demonstrated in bottlenose and roughtooth dolphins; scores on a composite 'Curiosity' rating factor was positively correlated with the time dolphins spent looking at a novel, surprising stimulus located outside their enclosure (Lilley, de Vere, Yeater & Kuczaj, 2018).

Outside of bottlenose dolphins, there are several studies which have demonstrated consistent individual differences in specific behaviors of wild grey seals. The time that dominant, resident males spent alert was highly consistent across two breeding seasons (Twiss & Franklin, 2010), while pup-checking behaviors performed by females and aggressive behaviors by males were consistent across an undisturbed and experimentally disturbed setting (Twiss, Culloch & Pomeroy, 2011). In the following year, female pup-checking frequencies showed reasonable stability, but there were individual differences in the extent of this consistency (Twiss, Cairns, Culloch, Richards, & Pomeroy, 2012). Finally, newly weaned grey seal pups showed individual differences in their aggressive, affiliative and checking behaviors during interactions with both familiar and unfamiliar pups (Robinson et al., 2015).

To date, there are only two assessments of broad personality dimensions in any pinniped species, each using either behavioral coding or trait rating. Analysis of the behavior of a captive California sea lion and harbor seal population revealed two reliable factors: Boldness and Routine Activity (de Vere, Lilley & Highfill, 2017). Boldness was interpreted as broadly analogous to the human factor of Extraversion, but also included behaviors indicative of the Openness-related traits of curiosity and exploration. Routine Activity was tentatively interpreted as exhibiting some potential parallels with Conscientiousness, due to the loadings of predictable, routine behaviors. Across species, there was substantial similarity in the content of these factors, suggesting some evolutionarily conserved traits. California sea lions housed across five facilities were also recently assessed for personality using trait rating (Ciardelli, Weiss, Powell & Reiss, 2017). Three factors emerged, labelled ‘Extraversion/Impulsivity’, ‘Dominance/Confidence’, and ‘Reactivity/Undependability’. Traits loading strongly on the first of these factors included playful and curious, thus demonstrating similarities with Extraversion and Openness dimensions, as well as items suggesting attention-seeking tendencies. The Dominance/Confidence factor showed similarities to the Dominance factor found in other non-human species, but lacked aggressive traits. Finally, Reactivity/Undependability contained several traits referring to interactions with humans, as well as items consistent with low Agreeableness.

These studies therefore provide evidence for broad personality dimensions across multiple populations of bottlenose dolphins, through both trait rating and behavioral coding, and for some specific traits and dimensions in a few populations of pinnipeds. However, attempts to validate broad personality dimensions across methods is lacking in marine mammals, and is still generally uncommon in animal personality research. Pinnipeds also lack the progress made in the bottlenose dolphin literature, particularly in that only one trait rating assessment has been conducted. Such research would facilitate further cross-species comparisons, as well as informing our knowledge of the validity of trait rating and behavioral coding when applied to marine mammals.

Dominance and Animal Personality

In humans, social dominance-related aspects of personality tend to be combined with non-dominance facets within a broader factor, such as Extraversion (DeYoung et al., 2007), or as a combination of traits from more than one factor, such as high Extraversion and low Agreeableness (Mehrabian, 1996). However, in the non-human personality literature, a Dominance factor has frequently emerged as distinct from other groups of traits. This factor tends to correlate significantly with dominance rankings (Gosling & John, 1999), and contains items such as ‘fearful’, ‘bullying’, ‘jealous’, and ‘independent’ (e.g. King & Figueredo, 1997). However, it is possible that this factor is not as ubiquitous as it initially appears, as in certain cases, the separable Dominance factor may be an artifact of specific methodology. In some studies, Dominance factors are derived from trait ratings that contain the trait words *dominant* and *submissive*, (e.g. Ciardelli et al., 2017; King & Figueredo, 1997), or other constituent traits that are defined in terms of dominance and/or submissiveness. It is possible that these aspects of the rating method contribute to two sources of ambiguity. First, they conflate the concepts of an individual’s social position in a dominance hierarchy, and the personality traits that the individual possesses; an animal may behave in a certain way because of their social rank, or because of their personality, or some combination of the two, but these individual characteristics are not equivalent. Second, these methods may obscure issues of causality; for example, if a low-ranking individual performs submissive behaviors, she may be performing these behaviors because she is in a low rank position, or these behaviors are characteristic of her personality and facilitated her obtaining a low rank. Patterns of behavior will certainly be determined by interactions between several factors, including social rank and personality, but any attempt to discern the contribution of single variables is made significantly more difficult if trait items that refer to ‘dominance’ are used.

It would therefore be informative to determine whether personality dimensions similar to the ‘Dominance’ factors found in some species are still consistently produced if the *dominant* and *submissive* trait words, as well as related wordings, are not used in ratings. Furthermore, if a similar factor does continue to emerge, the absence of these trait words may weaken the correlation between Dominance factor scores and dominance rankings. Even if a separate Dominance factor did not emerge under such methodology, it would still be informative to determine which other personality dimensions, or more specific facets, are related to dominance rankings. Expansion of the study of dominance and personality to other taxa may therefore shed light on whether similarly closely related species with different social structures and dominance hierarchies exhibit comparable personality-dominance associations.

Change in Animal Personality with Age

Although one of the defining features of personality is that individuals show consistency across time, in humans there are some known patterns of expected change. Across our lifespan, changes in average scores on certain traits and dimensions are seen until at least middle age. For example, from the mid-twenties until around 50 years old, Conscientiousness scores tend to increase, while Neuroticism scores decrease (Roberts, Walton & Viechtbauer, 2006). In the study of non-human personality, the potential for species-specific developmental trajectories has also been increasingly recognized.

Current evidence suggests that general behavioral repeatability does tend to decrease with increased inter-test interval (Bell, Hankison & Laskowski, 2009), as would be the case for humans. While some of this variation can likely be attributed to a range of situational confounding variables and measurement bias, a portion may reflect the presence of age-related changes in personality. Subject populations often contain animals of various ages, but age is typically not accounted for in statistical analyses (e.g. King & Figueredo, 1997); this may therefore obscure possible developmental changes in personality. In other studies, the possibility of age effects is taken into account (Sussman, Bentson & Crockett, 2013).

In those non-human personality studies which do consider the potential role of age, various patterns have emerged. One common theme is a decrease in scores on both Extraversion-like and Openness-like factors. For example, gorilla scores on Agreeableness, Sociability, and Openness were lower for older individuals, with a more dramatic change in males for Sociability (Eckardt et al., 2015). Both domestic cat and snow leopard adults have also demonstrated lower Openness and Extraversion scores compared to younger animals (Gartner, Powell & Weiss, 2014). Similarly, adult California sea lions were rated lower on the Extraversion/Impulsivity factor than individuals under five years old (Ciardelli et al., 2017). The non-human Dominance factor has also shown developmental changes in several species, such as increases with age in chimpanzees (Weiss et al., 2009) and orangutans (Weiss & King, 2015).

Other studies have found entirely different personality dimensions for animals of different ages. Comparisons of studies assessing chimpanzee populations of different ages tentatively suggest that infant chimpanzees exhibit a separate Activity dimension which does not emerge in adults, a developmental pattern also seen in humans (Gosling & John, 1999). In rhesus macaques, only the Fearful dimension emerged consistently between the ages of one and seven, while Aggression emerged clearly only after the third year (von Borell, Kulik & Widdig, 2016). It is difficult and time consuming to conduct longitudinal studies of non-humans, which would be the most accurate method to determine whether developmental changes in personality exist. However, given the current evidence for some age-related differences, it is worth incorporating the age of subject animals into studies of non-human personality.

Emotions and Animal Personality

As in the human emotion literature, debate exists over the definition of animal emotions. Here, the term emotion will be used to refer to all of the experiential mental states that have been labelled emotion, affect, feelings, and moods (de Vere & Kuczaj, 2016), but ‘affect’ may be used in some instances for the sake of consistency with existing models. Emotions are relevant to the study of personality because there is substantial overlap between the terminology used in both literatures in humans (Izard, Libero, Putnam & Haynes, 1993). For example, consider the Positive Affect and Negative Affect Model, a widely accepted description of human emotional experiences (Watson & Tellegen, 1985). High Positive Affect is characterized by states such as enthusiasm, excitement and activity, whilst high Negative Affect is characterized by distress, fear and hostility. In some literature (Paul, Harding & Mendl, 2005), the Valence-Arousal model is preferentially used, which simply consists of the statistically unrotated Positive-Negative Affect model. In this alternative, Valence is characterized by pleasantness versus unpleasantness, and Arousal describes the degree of arousal or activation (Russell, 2003). These latter dimensions tend to be used more in animal research, as they are more easily isolated using behavioral measures, in the absence of self-report data (Paul, et al., 2005).

Some draw a distinction between trait affect, a person's individual propensity to experience a specific emotional state, and state affect, that person's capacity to experience the state at all (Izard et al., 1993). Trait affect contributes to several of the 'big five' human personality dimensions, as many personality traits have emotional content. For example, Neuroticism has been correlated with Negative Affect, and Positive Affect with Extraversion (Costa & McCrae, 1980; Tellegen, 1985). A large number of the items found in Extraversion and Openness directly refer to emotions (e.g. excitement, curious: Pytlik Zillig, Hemenover & Dienstbier, 2002), and other factors still contain certain traits referring to emotional states, such as trait anger as a marker of low Agreeableness (Kuppens, 2005). An emotional response to experiences of beauty is also a reliable cross-cultural indicator of Openness to Experience (McCrae, 2007). In the context of personality research, trait words with emotional content therefore refer to trait affect. Unless these trait words are operationally defined in a way that excludes emotional content, it seems likely that raters may interpret these items as they are typically used. This may cause implicit assumptions to be made about the capability of a studied species to experience these emotions. While there is substantial evidence for the shared neurological features underlying basic emotional mechanisms across species (e.g. Panksepp, 2011; Paul et al., 2005), little work has focused on the potential for species-specific emotional repertoires. This is somewhat surprising, given the attention paid by many studies to the species-specific nature of personality traits (e.g. Freeman et al., 2013). Thus, it is both interesting and significant to consider how best to determine whether emotional trait words have the same relevance and meaning for non-human species.

It is currently not practical to use experimental tests to determine a species' capacity for particular emotions. While physiological and neurological methods to study emotional states are being developed, they tend to have several significant associated problems. For example, many of these tests involve invasive procedures that are likely to impact the animal's current state (Broom, 1993), therefore producing inaccurate information. Given that the same levels of valence and arousal may be associated with different emotional states, even known physiological correlates tend not to be specific to a single state (Broom, 1993; Dawkins, 2001). Behavioral indicators have fewer associated logistical challenges, but they still suffer from the problem of understanding which indicator is associated with which particular state. Furthermore, using such experiments to test animals for a range of emotions is not currently a time efficient data collection strategy.

As trait ratings of animal personality have been shown to be reliable and broadly valid (according to correspondence with behavioral measures), raters seem to be able to detect observable behaviors associated with particular traits. They may therefore also be able to reliably rate whether a species appears able to experience certain emotional states, again based on external indicators. This is likely to be particularly true for social species and/or emotional states, for which the presence of observable indicators of emotions would be extremely advantageous in interactions with conspecifics (Buck, 1999; Kuczaj, Highfill, Makecha & Byerly, 2013). Personality traits with emotional content could then be selected on the basis of whether the relevant emotional state has been identified as experienced by the species under study. This approach would share some parallels with the existing ‘expert’ selection method used in animal personality research. ‘Experts’, individuals who are extremely familiar with a species, nominate trait items which are considered relevant to the focal species (Ciardelli et al., 2017; Gosling, 1998; King & Figueredo, 1997). Given the crossover of personality and emotional terminology, nominated trait words often include those with emotional content. Asking people who have experience with a species of interest to rate that species’ ability to experience a specific emotional state could therefore be seen as a formal extension of this existing ‘expert’ methodology. Additionally, one study has assessed owners for their perceptions of the emotions experienced by their pets, with owners giving ‘yes’ or ‘no’ responses to questions “is your animal ever *emotion*?” (Morris, Doe & Godsell, 2008). Embarrassment was reported with the lowest frequency, but almost 20% of 907 pets were still rated as having experienced this state, while fear and interest were reported in almost 100% of rated animals. Furthermore, raters generally reported high confidence in their judgements. This therefore provides evidence that pet owners feel able to rate familiar, individual, domestic animals on the emotions they experience. However, whether raters can detect and use external indicators of emotional states in non-domestic species to produce reliable ratings of the emotions experienced by the species as a whole has yet to be explicitly assessed.

Current Study

In summary, personality research in pinniped species is limited. No multiple method studies of personality exist for this taxon, and examinations of potential personality-dominance relationships are minimal. Furthermore, as discussed above, any studies of animal personality should consider the age of studied individuals. The potential for emotional trait words to be more (or less) relevant for non-human species has also not been addressed. The current study addresses these issues by extending a previous behavioral coding assessment of two pinniped species (de Vere, Lilley & Highfill, 2017), California sea lions (*Zalophus californianus*) and harbor seals (*Phoca vitulina*), using the trait rating method. The predictions for the current study were as follows:

- Personality trait ratings were expected to be largely reliable. Two factors were found in the behavioral coding assessment of these species (de Vere et al., 2017) and most species exhibit between two and four personality factors (Gosling & John, 1999), so between two and four trait rating factors were expected to emerge in each pinniped species. Consistent with life history characteristics and previous assessment (de Vere et al., 2017), California sea lion factors were expected to contain more high-energy and social traits than those of harbor seals.
- Both the length of experience that raters had with each animal and the animal's age were expected to correlate negatively with the extent to which raters were unsure about their answers, as more exposure to an animal should theoretically provide a rater with more data upon which to base their ratings.
- The current study aimed to provide the first assessment of methodological convergent validity for pinniped personality. Some convergence is expected to be found, but consistent with previous research (e.g. Barnard et al., 2016), factors produced from one method were expected to show moderate to strong correlations

with more than one factor produced from the other method, and/or show otherwise imperfect correspondence.

- Ratings of dominance rankings were predicted to correlate moderately with at least one behavioral and/or trait rating factor, which was expected to contain traits comparable to other non-human Dominance personality factors (e.g. confident, jealous). Larger correlations were expected to be seen in California sea lions, due to their stronger dominance hierarchies (Riedman, 1991).
- Non-humans share many aspects of personality with humans, and thus are also likely to show changes in personality with age. This is particularly true for social species, for whom it is likely adaptive to adopt personality traits conducive to reproduction, parenting, and higher positions in social hierarchies in adulthood. As seen in humans (Roberts et al., 2006), older animals were expected to score higher on factors containing behaviors or traits associated with Conscientiousness. Based on the existing non-human literature, species-specific patterns were also expected, such as decreases in Extraversion- and Openness-related traits (Eckardt et al., 2015; Gartner, Powell & Weiss, 2014) and increases in aggression/dominance related traits with age (von Borell et al., 2016).
- Finally, ratings of the emotions experienced by both species were expected to be reliable for at least some states, and to show broad consistency across species.

CHAPTER II – MATERIALS & METHODS

Subjects

Animal subjects were the pinniped population at Six Flags Discovery Kingdom, Vallejo CA (Table 1). The population includes individuals of two species: eleven California sea lions (*Zalophus californianus*) (6 male, mean age 8.4 years) and seven harbor seals (*Phoca vitulina*) (2 male, mean age 5.5 years).

Materials

First, human perceptions of seal and sea lion emotional repertoires were assessed, in order to create personality questionnaires for each species. To design the emotional repertoire questionnaire, a list of emotional state items was generated. All items from a common human assessment of affect, the PANAS scale (Watson et al., 1988), were included, with one exception; *strong* was replaced with another word from the same content category, *confident* (Zevon & Tellegen, 1982), due to possible confusion of the mental state *strong* with a physical attribute, as well as the use of confident in previous personality studies. An extensive analysis of affective terminology (Ortony et al., 1987) was then cross-referenced with trait words used in the study of bottlenose dolphin (Highfill & Kuczaj, 2007), and chimpanzee (King & Figueredo, 1997) personality to identify which personality traits contain emotional tendencies. These two personality studies were selected to generate the item pool due to their relevance in assessing the two species in the present study; King and Figueredo's (1997) assessment was one of the earliest studies to apply the trait rating method to animals, and provided the basis for standard rating assessments of primate personality. Similarly, Highfill and Kuczaj's (2007) study was, at the time, the only broad trait rating assessment of any marine mammal species, and therefore was likely to contain traits relevant to the relatively phylogenetically similar pinniped species. Any trait words implying emotional content that were not already included from the PANAS were selected. This produced a list of 46 trait items: 20 from the PANAS scale, one from Openness to Experience, three from Conscientiousness, one from Dominance, five from Extraversion, nine from Agreeableness, and seven from Neuroticism (Table 3).

These items were assembled into one questionnaire for each species, both of which contained the same 46 items. Each state was framed in the question: to what extent do you think that 'species' are capable of experiencing 'x' state'? (e.g.: 'To what extent to you think that harbor seals are capable of experiencing fear?' or 'To what extent do you think that California sea lions are capable of experiencing fear?') Each question was followed by a 7-point Likert scale (1 = *is not capable of experiencing the state*; 7 = *appears completely able to experience the state and does so often*). An 'unsure' option was also included so that raters would not feel forced to make a judgement if they were uncertain. Instructions preceded each questionnaire, asking the rater to base their ratings on their general experience with the species, not necessarily just the individual animals under study here. The order of state items was randomized for each rater, in order to minimize possible rater fatigue and order effects.

Five raters completed this questionnaire for each species, all of whom were female pinniped trainers at Six Flags Discovery Kingdom, Vallejo (Table 2). One rater had only two weeks of experience with harbor seals (specifically, the subjects of this study) upon beginning their ratings, so their data were not included in the harbor seal analysis. Overall rater agreement for each species' questionnaire was good, as assessed by intra-class correlation coefficients [California sea lions: ICC(3,k) = 0.861, harbor seals: ICC(3,k) = 0.822]. However, it was not possible to compute intra-class correlation coefficients, or other standard reliability measures, for each individual state due to the absence of repeated ratings for each state (one from each rater) for each rated state. As a proxy reliability measure, a conservative cut-off of one standard deviation across raters was used to identify unreliably rated states, which in practice reflected a maximum difference of two points between raters. This resulted in a pool of 21 reliable items for California sea lions and 25 for harbor seals (with 14 shared across species; Table 3). As a rating of three indicated that raters believed the species may be able to experience the state but do so negligibly, any reliable items with an average rating of three or lower were removed. For both species, this resulted in the removal of *inspired*, *ashamed*, and *jealous*, with harbor seals additionally rated as incapable of experiencing *guilty* and *sympathy*. This resulted in 18 remaining states for California sea lions and 20 for harbor seals. The 11 states that might be considered in some way complex or secondary were rated with slightly greater variability, with an average standard deviation of 1.53, while ratings of the 35 remaining primary emotional states had a standard deviation of 1.15.

All reliable state items with averages above three were added to the pool of potential traits for personality questionnaires. Non-emotional traits from studies of chimpanzee (King & Figueredo, 1997) and bottlenose dolphin (Highfill & Kuczaj, 2007) personality were also added to this pool of potential traits, excluding *dominant* and *submissive*. From this pool, 30 items were selected for each species' personality questionnaire, such that all human Five Factors, plus chimpanzee Dominance (King & Figueredo, 1997) were represented, with minimal redundancy (Table 4). These 30 items were assembled into two, species-specific questionnaires. Each questionnaire was preceded with instructions similar to those preceding the emotional questionnaires, instructing raters not to discuss their ratings with anyone else, and explaining the rating scale. Each questionnaire also asked raters to give their months/years of experience with the rated animal, and to place them into a dominance ranking category: low, medium, high.

Procedure

As raters would be completing a relatively large number of questionnaires (one for each of 18 animals), several approaches were taken to minimize rater fatigue. Firstly, raters received batches of a maximum of six questionnaires at one time; the identity of the animals included in each set were randomized. Once the rater had completed and returned all six, they were then provided with the next set. Secondly, the order of traits was randomized for each questionnaire received within the same batch. Finally, the animal identities assessed by each set of questionnaires was randomized for each trainer. This latter approach also ensured that questionnaires were completed for all subjects, even though some raters did not complete all 18 questionnaires. All raters were SFDK pinniped trainers who had previously completed the emotional state questionnaires. Due to changes in staffing over the course of data collection, some raters only completed a subset of questionnaires, and one rater did not have any experience with three individuals. However, between two and four raters completed questionnaires for each animal, between August and December 2016.

Data Analysis

All analyses described below were conducted using IMB SPSS 22.0 for Windows and/or MATLAB.

Trait Rating

Inter-rater reliability was assessed for each trait, in each species, using intra-class correlation coefficients (ICCs). Two types of ICC were calculated; ICC(3,1), to assess the reliability of an individual measurement, and ICC(3,k), to assess the reliability of k ratings of a trait (where k equals the number of ratings). Missing values were replaced by the average rating for each trait (Morton et al., 2013), to facilitate the calculation of intra-class correlation coefficients without artificially increasing reliability values. As an intra-class correlation coefficient greater than zero indicates above chance agreement between raters (Freeman et al., 2013), all traits with positive coefficients were considered reliable and included in further analyses.

Reliably rated traits were analyzed using a regularized exploratory factor analysis (REFA) for each species. This statistical method has proven superior to principal components analysis (PCA) when sample sizes are small (i.e. below 50) (Jung & Lee, 2011), and additionally provides consistency with the only other trait rating assessment of California sea lions (Ciardelli et al., 2017). As in this recent study, a PCA was also conducted for each species for comparison to REFA results. The number of components or factors to extract for each analysis were determined using a scree plot and parallel analysis (Horn, 1965). An oblique rotation method was then applied to the components/factors extracted, to allow for possible non-independence of psychological constructs. Trait loadings greater than 0.35 were retained, and in the case of crossloadings, a trait was considered to 'belong' to the dimension on which it had the highest loading. Items with no loadings above 0.35 in the initial REFA solution were removed sequentially until all traits had significant loadings. Cronbach's alpha was calculated for each factor to assess internal consistency.

Rater Experience

To assess the potential impact of rater-animal experience on trait ratings, Spearman's rank-order correlation coefficients were calculated between raters' months of experience with an animal, and the number of unsure and blank responses given for their rating of that animal. Correlations were also calculated between each animal's age and the number of unsure and blank responses. In both cases, a correlation coefficient was calculated separately for each species.

Dominance

As every animal was not rated by each rater, percentage agreement was calculated for dominance rankings (low = 1, medium = 2, high = 3). Intra-class correlation coefficients were also calculated, with missing values replaced by the animal's average dominance ranking.

Several Spearman's correlation coefficients were calculated, between dominance rankings and factor scores on each elucidated personality factor, separately for each species. Behavioral coding factor scores have been calculated previously (de Vere, Lilley & Highfill, 2017), and trait rating factor scores were calculated by multiplying an individual's average score on a trait with that trait's factor loading, and then summing these products for all traits on a factor.

Personality & Age

To assess whether any association exists in the subject population between animal age and personality factor scores, several Spearman's correlation coefficients were calculated. A correlation coefficient was calculated between scores on each trait rating and behavioral coding factor in each species, and animal age. For significant correlations, age and factor scores were graphed to visually examine the linearity of patterns of change.

Convergent Validity

The convergent validity of the trait rating and behavioral coding methods was determined using Spearman's correlation coefficients between animals' factor scores across methods. Given the mixed evidence for good cross-method convergence of broad animal personality dimensions, a number of correlations were also calculated between several behaviors and traits according to a priori predictions. This was intended to provide confirmation (or not) of more specific convergent validity, and to determine whether, in the case of poor factor correspondence, discrepancies were attributable to the lack of convergence of the broad factors or simply to an overall lack of methodological convergence. The behaviors and traits predicted to be correlated for each species were: resting and active, alert and scanning, and alone play/social play and playful. No measures were employed to control for family-wise error, despite the large number of correlation coefficients calculated, because this would virtually eliminate any possible significant or meaningful findings, particularly given the already limited statistical power.

CHAPTER III - RESULTS

Trait Rating

Elucidated trait rating factor solutions are described, interpreted, and labelled below. Labels for these factors were intended to provide the most informative and summative description of both the positive and negative pole of each factor (PositiveLabel/NegativeLabel), but in cases where a factor had only one item or no items loaded at one of the poles, the factor was labelled based solely on the dominant pole.

Harbor Seals

On average, raters had 1.73 years of experience with the rated animals (standard deviation = 2.12 years). Only "Unexcitable" had an ICC(3,1) estimate of zero or less for harbor seals, and was therefore excluded from all further analyses. The remaining ICC(3,1) values averaged 0.269 and ranged from 0.071 for "Quiet" to 0.553 for "Playful". The average seal ICC(3,k) estimate was 0.560, and ranged from 0.234 for "Quiet" to 0.832 for "Playful".

Both a scree plot and parallel analysis suggested that three factors be extracted. Initial factor loadings are shown in Table 4. Traits with loadings of less than 0.35 on any factor were removed one by one (lazy, simple, unoriginal, solitary) until all remaining items had loadings greater than 0.35. The final REFA factor structure is shown in table 5. These three factors explained 15.62% of the total variation (Table 5), as REFA variance estimates are inherently low compared to PCA estimates (e.g. 16.04% and 54.88% for the initial REFA and PCA solutions, respectively). Cronbach's alpha was calculated for each factor as a measure of internal consistency, both for the initial and final REFAs. All alpha values exceeded 0.7, indicating acceptable internal reliability (Nunnally, 1978; Tables 4

& 5). Despite the rotation method allowing for inter-factor correlations, these were low, with a maximum of 0.162.

Factor 1 in the final REFA solution had 15 traits with loadings greater than 0.35, 11 of which had their highest loadings on this factor and were therefore counted as constituting the factor. Positively loaded items were: *curious*, *demanding*, *stable*, *enthusiastic*, and *interested*, while negatively loaded items were: *cautious*, *sensitive*, *fearful*, *dependent*, *jittery*, and *nervous* (Table 5). Several of the traits which loaded onto the negative pole can be found in human Neuroticism, including *nervous, fearful,* and *dependent* (Goldberg, 1990), as well as other items consistent with the content of Neuroticism, such as *jittery* and *sensitive*. Openness-like tendencies present at the positive pole of Factor 1, as indicated by items *curious* and *interested*, are combined with traits suggesting further information- and attention-seeking tendencies. This factor was therefore labelled Interest/Neuroticism.

Eight of the ten traits with significant loadings on Factor 2 loaded most strongly on this factor. *Stubborn*, *incompliant*, *temperamental* and *bullying* loaded positively, while *quiet*, *gentle*, *tolerant*, and *calm* loaded negatively on Factor 2. It therefore contains a number of items indicative of instability and impulsivity at the positive pole (i.e. *stubborn, bullying, temperamental,* and *incompliant*), which resemble a combination of high Neuroticism and low Agreeableness (Goldberg, 1990). At the negative pole, loaded traits quite strongly resemble high Agreeableness, with items such as *gentle* and *tolerant*. Given the combination of tendencies at the positive pole, this factor was labelled based on the strongest loaded item at this pole and the overall tendencies at the negative pole: Stubborn/Agreeable.

On Factor 3, six of the 11 items with loadings greater than 0.35 loaded highest on this factor. Five of these items had positive loadings: *intelligent*, *inquisitive*, *alert*, *active*, and *playful*, while predictable loaded negatively. The traits clustered on Factor 3 are almost exclusively items from human Extraversion (*playful* and *active*) and Openness (*intelligent*, *inquisitive*, and *alert*) (Goldberg, 1990). The only exception to this is the single negatively loaded item, *predictable*. If this item reflects conventionality, this is theoretically consistent as an opposing pole to intelligent and playful tendencies. This factor was therefore labelled Extraversion-Openness, on the basis of the dominant positive pole.

California Sea Lions

For California sea lions, all 30 traits had ICC(3,1) estimates greater than zero, with an average of 0.36 and ranging from 0.11 for "Curious" to 0.64 for "Stable". Sea lion ICC(3,k) estimates ranged from 0.32 to 0.88 for the same traits, respectively, with an average of 0.67.

Both a scree plot and parallel analysis suggested that three factors be extracted. Initial factor loadings are shown in Table 7. Traits with loadings of less than 0.35 on any factor were removed one by one (simple, intelligent, protective) until all remaining items had loadings greater than 0.35. The final REFA factor structure is shown in Table 8. These three factors explained 17.84% of the total variation (Table 8). Cronbach's alpha values for each factor, both for the initial and final REFAs, indicated adequate internal consistency, with all values approximately 0.7 or greater (Table 7; 8). Only the correlation between factors two and three was appreciable, at -0.3941.

In the final REFA solution, Factor 1 had 16 traits with significant loadings, of which 13 were highest on this factor. Loaded positively were: *predictable*, *quiet*, *cooperative*, *calm*, and *stable*, and loaded negatively were: *alert*, *aggressive*, *incompliant*, *demanding*, *temperamental, anxious, bullying*, and *irritable* (Table 8). This factor shares three traits directly with the previously elucidated sea lion Reactivity/Undependability factor (Ciardelli et al., 2017): *cooperative, irritable* and *aggressive,* although *aggressive* in the R/U factor is specifically directed towards people. Factor 1 contained several additional traits which were consistent with stability and predictability at the positive pole (i.e. *stable, calm, predictable*), and low agreeableness and some neuroticism-like tendencies at the negative pole (i.e. *incompliant, irritable, bullying*) (Goldberg, 1990). This factor was therefore labelled Agreeable/Incompliant.

Ten of the 13 items with significant loadings on Factor 2 loaded highest on this factor. Curious, *inquisitive, active*, *playful*, *energetic*, *excitable*, and *enthusiastic* loaded positively, with *unoriginal*, *lazy*, and *solitary* loaded negatively. *Enthusiastic*, *curious*, and *playful* traits loaded on both this factor and the Extraversion/Impulsivity factor (Ciardelli et al., 2017), while these factors also had additional traits with similar meanings (i.e. *inquisitive* and *creative*). However, unlike the Extraversion/Impulsivity factor, Factor 2 had several negatively loaded traits indicating inactive and solitary tendencies. Additionally, Extraversion/Impulsivity has a further attention-seeking element characterized by the trait words: *demanding*, *impulsive*, *jealous*, and *aggressive* to sea lions, which Factor 2 is lacking. Given these similarities, it is therefore consistent that four of the items loaded on Factor 2 (*enthusiastic*, *playful*, *energetic*, *active*) can be found in the original human Extraversion factor, and three in human Openness to

Experience (*curious*, *inquisitive*, *unoriginal*) (Goldberg, 1990). Overall, Factor 2 contains many items which have previously formed part of personality dimensions homologous to Extraversion, in humans (Goldberg, 1990), chimpanzees (King & Figueredo, 1997), as well as the only other trait rating assessment of this species (Ciardelli et al., 2017). Thus, Factor 2 was labelled Extraversion/Openness.

On the third factor, only four of the ten significantly loaded items had their highest loading on this factor. Items with negative loadings were: *fearful*, *dependent*, and *nervous*, while only *determined* loaded positively. This factor shows some parallels with human Neuroticism (Goldberg, 1990), due to the loading of items *fearful*, *nervous*, and *dependent*, and also has some similarities with the previously elucidated Dominance/Confidence sea lion factor (Ciardelli et al., 2017), although these factors are not as directly analogous as the other comparisons discussed above. As only one trait, *determined*, characterized the positive pole of Factor 3, this factor was labelled 'Dependency' based on the negative pole of the dimension.

Rater Experience

For harbor seals, raters chose the uncertain option or left the question blank on 40 of 750 total ratings made of individual traits, comprising 5.33%. The same was true of 8.96% of California sea lions ratings, reflecting 86 uncertain responses of 960 total ratings.

Neither correlation for harbor seals was significant, either for rater uncertainty with the length of experience with the rated animals ($r = -0.189$, $p=0.16$) or with animal age ($r = -0.185$, $p = 0.168$).

For California sea lions, the correlation between rater uncertainty and length of experience was not significant (r = -0.269, p = 0.136), but there was a significant negative correlation between uncertainty and animal age (r = -0.401, p = 0.023).

Dominance

No personality factors were significantly correlated with harbor seal dominance rankings. However, two correlations were approaching significance, with trait rating factor 3 (r = -0.721, p= 0.068; Figure 3a) and with Boldness (r = -0.685, p = 0.09; Figure 3b) (Table 10). The equivalent California sea lion factors, trait rating factor 3 and Boldness, were significantly correlated with dominance rankings (r = 0.799, p = 0.003; r = -0.851, p = 0.001) (Figure 4, Table 11).

Personality & Age

Harbor seal age was significantly negatively correlated with the Boldness personality dimension (r = -0.955, p = 0.0.01), indicating that Boldness scores decrease with increasing age (Table 10; Figure 5). California sea lion age was also significantly negative correlated with both trait rating factor 3 (r = 0.825, p = 0.002; Figure 6a) and Boldness (r = -0.793, p = 0.004; Table 11; Figure 6b).

Age and dominance were also significantly positively correlated in both species (harbor seals: r = 0.809, p = 0.028, California sea lions: r = 0.941, p<0.001) (Figures 7 & 8).

Convergent Validity

There were no significant correlations between harbor seal trait rating and behavioral coding personality dimensions (Table 12).

California sea lion Boldness and trait rating factor 3 were significantly correlated ($r = -0.727$, $p = 0.011$) (Table 13). The correlation between Routine Activity and trait rating factor 2 was also approaching significance ($r = 0.582$, $p = 0.06$).

Convergence of individual traits and behaviors was mixed. For harbor seals, two of the five predicted correlations were significant or approaching significance, between Playful and alone play and between Active and resting, respectively (Table 14). However, there were several other unpredicted significant correlations: significant positive correlations between Active and alone play and between Curious and scanning, and a negative correlation between Playful and resting which was approaching significance (Table 14). A similar pattern emerged for California sea lions, with the same two predicted trait-behavior pairs showing significant or approaching significant correlations (Table 15). Additionally, there were significant positive correlations between Playful and scanning and Playful and tactile, a significant negative correlation between Active and resting, and a negative correlation between Alert and resting which was approaching significance (Table 15).

CHAPTER IV – DISCUSSION

Emotional Repertoires

The present study provides the first direct assessment of the ability of humans to rate the emotional states experienced by non-humans at the species level. The reliably rated emotional states were highly similar across species, likely reflecting both the comparable contexts in which raters had experience with each species (i.e. currently at the same facility) and the close phylogenetic relationship between California sea lions and harbor seals (Arnason et al., 2006). However, there were still some notable differences. For example, sea lions but not seals were rated reliably on the following states: *determined, excitable, energetic, irritable, protective, cooperative,* and *aggressive,* while seals but not sea lions were rated reliably on: *interested, afraid, jittery, guilty, cautious, sensitive, gentle, stubborn, sympathetic, tolerant,* and *unexcitable* (Table 3). Those states apparently 'unique' to sea lions overall reflect high energy (e.g. *excitable* and *energetic*) and/or social emotions (i.e. *protective, cooperative,* and *aggressive*), and this species was also rated higher on *enthusiastic* than were seals. California sea lions are highly social, and have frequently been described as 'gregarious' and 'friendly' (Heath & Perrin, 2009). During the breeding season, groups of up to approximately 20 female sea lions form "milling" aggregations prior to mating, and males fight to obtain territories which are maintained through this season. The sea lion behavioral repertoire also contains a number of high energy behaviors, such as leaping and various types of play (Riedman, 1991). It is therefore consistent with these behavioral and life history characteristics that trainers rated California sea lions as capable of experiencing both high arousal and social emotional states. In contrast, harbor seals have been described as easily disturbed and highly vigilant (Schusterman, 1968; Stevens, Thyssen, Laevens & Vervaecke, 2013). This is consistent with the raters' view that this species experiences states such as *jittery* and *cautious*. They are also pseudo-social, in that although they frequently gather in huge numbers at haul out sites, they remain several body lengths apart from one another, do not show stable social groupings, and tend to engage in social interactions only for breeding and agonistic purposes (Bigg, 1981; Godsell, 1988). That trainers did not reliably rate them as capable of experiencing the same social emotions as California sea lions may therefore be related to these species' life history and behavioral tendencies.

Raters agreed that three states are not experienced by either pinniped species (*jealous, ashamed,* and *inspired*), with harbor seals additionally incapable of experiencing *guilt* and *sympathy*. It is interesting that four of these states (*ashamed, guilty, sympathetic,* and *jealous*) can be classed as complex or secondary emotions (Johnson-Laird & Oatley, 1989), in that they rely on self-consciousness or self-awareness (Lewis, 2002) and may additionally involve evaluative appraisals relative to one's own goals or knowledge (Levine, 1997; Morris, Doe & Godsell, 2008). In fact, the only emotional item that could be considered in any way secondary that was rated reliably as being experienced by one of the species was *protective*; this state implies a self-other distinction, and therefore requires some level of self-awareness. All other secondary emotional states were either not rated reliably or rated as not experienced by either species. Because of these cognitive requirements, the ability to experience this type of complex emotional state is often attributed only to humans and possibly some non-human primates (Morris, Doe & Godsell, 2008; Preston & de Waal, 2002). However, it has been suggested that dolphins may be capable of the self-consciousness required for secondary emotions (Marten & Psarakos, 1995; Morris, Doe & Godsell, 2008). There is also an evolutionary argument for social species to be capable of experiencing social and moral emotions (e.g. *jealousy* and *contempt*, respectively; Buck, 1999), as these are thought to be useful in guiding and interpreting social interactions with conspecifics in much the same way as in humans (Kuczaj et al., 2013). Contrary to these arguments, raters in the current study did not believe that either California sea lions or harbor seals were capable of experiencing these four secondary or complex emotions. However, this is consistent with the previous finding that raters may be more confident when attributing primary emotions to rated animals than secondary emotions (Morris, Doe & Godsell, 2008). Additionally, there were other secondary emotions included in the full questionnaire which were not rated reliably (i.e. standard deviations greater than one), such as *proud, greedy*, and *selfish*. Ratings by individual trainers for these states ranged from one (indicating '*are not capable of experiencing the state'*) to seven (indicating '*appear completely able to experience the state, and do so often*'), in some cases as extreme as two trainers rating the species at one or two, while the other trainers rated them a six or seven; some secondary emotions therefore seemed to polarize raters in their opinion. Given this, it is possible that secondary emotions in these species have fewer, unobservable, or more variable external indicators, rather than that pinnipeds are in fact incapable of experiencing secondary emotions. Additionally, the contexts in which raters in this study have experience with California sea lions and harbor seals likely influenced their perceptions of emotional experiences. For instance, trainers typically interact with animals in positive settings, such as feeding sessions, which may reduce the likelihood of their observing cases in which animals are *jealous* or *selfish*, as all animals are provisioned and are not in conflict. It is also possible that in undomesticated animals, such as marine mammals, are more likely to experience such states when interacting with conspecifics, but not with humans. If this is the case, the likelihood of trainers observing animals experiencing secondary emotions would therefore be reduced. However, these explanations are speculative, and further research is required to determine whether any are valid.

Regardless of these possibilities, these results still unquestionably reflect the perceptions that raters had of non-human emotional states, and thus rater agreement indicates reliability rather than validity. However, this approach shares many similarities with methods used in previous assessments, which have inferred that these types of ratings do reflect internal characteristics, such as emotions. For example, multiple studies have determined that caregivers are able to produce reliable ratings of the subjective well-being of individual apes, via ratings of the animal's happiness, the pleasure they derive from social situations, the extent to which they experience positive and negative moods, and their success at achieving their own goals (King & Landau, 2003; Weiss, King & Hopkins, 2009; Weiss, King & Murray, 2011); this body of research has also recently been successfully extended to several felid species (Gartner, Powell & Weiss, 2016). Additionally, hundreds of pet owners have confidently rated the emotions experienced by their pets (Morris, Doe & Godsell, 2008). Evidence of the validity of similar assessments comes from pigs, in which subjective ratings accurately discriminated between individuals reared in different settings (Wemelsfelder, 1999).

There are also examples in the human literature, such as assessment of cognitive abilities in children via reports by their mothers (Bornstein, Giusti, Leach & Venuti, 2005). Therefore, while ratings in the present study undoubtedly reflect rater perceptions and were influenced by rater characteristics and the taxa under study, it is not unreasonable to infer that these ratings, to some extent, may reflect accurate interpretations of the subjective, valenced experiences of non-human animals.

Furthermore, many of the reliably rated state words in the present study were used in the subsequent personality assessment of individual California sea lions and harbor seals. Many of these ratings are significantly correlated with relevant behaviors performed by these individuals (de Vere & Levine, under review), thus suggesting that raters did use observable characteristics and tendencies to make their judgements. A subset of these correlations can be seen in Tables 14 and 15, as all four individual trait words examined for specific convergent validity in the present study have emotional content and were therefore part of the original emotional state questionnaires.

Relatively few emotional states were rated reliably; just 46% and 54% of the 46 states in sea lions and seals, respectively. However, because it was not possible to calculate statistical reliability measures for this data, the standard deviation cutoff for a state to be considered reliably rated was extremely conservative (one standard deviation). For comparison, all but one average intra-class correlation coefficient for the personality rating data was above zero, indicating greater than chance rater agreement, but ratings of only two traits had standard deviations of one or less across both species. In other words, had the conservative one standard deviation cutoff been applied to the personality data, only two traits would have been considered 'reliable', while traditional, statistical measures of reliability found ratings to agree at above chance levels for all but one trait. Like the emotion ratings, this variability is likely attributable to the small number of raters, as a single rating had the potential to greatly inflate a trait's standard deviation. However, this provides a good comparison for the similarly variable emotion ratings. For personality traits with 'good' intra-class correlation coefficients above 0.7, standard deviations range from 1.26 to 1.88 for sea lions and from 1.25 to 1.68 for seals. If these standard deviation upper limits were used as cutoffs for the emotional ratings, 12 more states would have been classed as reliably rated for seals, and 14 more for sea lions. It is therefore highly unlikely that this assessment of species' emotional repertoires overestimates the extent to which California sea lions and harbor seals experience emotional states, and if anything, likely provides an underestimate.

Trait Rating

Harbor Seals

Of the three factors elucidated, the strong nervousness element of the Interest/Neuroticism dimension is particularly relevant for harbor seals, as this species has been described as particularly vigilant and easily disturbed (Schusterman, 1978). Consistent with this, the individual seals in the present study broadly have negative scores on this factor, with the sole exception of seal seven (Figure 1), indicating that raters judged them to tend more towards the 'Neuroticism' pole of this factor, and thus are more *cautious*, *sensitive,* and *fearful* than *curious* or *demanding*. Scores on the Stubborn/Agreeable factor for the studied seals were also predominantly negative (Figure 1). Raters therefore judged these individuals to be more 'Agreeable' than 'Stubborn', tending more towards being *quiet*, *gentle* and *calm* than *stubborn*, *temperamental,* or *bullying*. Finally, given that the negative pole of Extraversion-Openness is comprised of only one trait, it is unsurprising that all individuals' factor scores are strongly positive, although some are rated as more *intelligent*, *active* and *playful* than others (Figure 1). These findings are broadly consistent with prior non-human personality literature, given the emergence of factors resembling human Extraversion, Agreeableness, Neuroticism, and Openness (Gosling & John, 1999).

California Sea Lions

Ratings of individual sea lions in the present study tend towards the positive pole of Agreeable/Incompliant, although four animals have factor scores close to or below zero (Figure 2). This demonstrates that most of the studied individuals were rated as being more *alert*, *aggressive*, and *incompliant* than *predictable* and *calm*. Individual sea lions predominantly had highly positive scores on Extraversion/Openness (Figure 2), indicating ratings tended more towards *intelligent*, *curious* and *active* than towards *unoriginal* and *lazy*. All average Dependency factor scores for sea lions in the present study were negative, although the standard error ranges for several individuals span positive scores (Figure 2). These sea lions were therefore rated as being more *quiet* and *fearful* than *protective* or *determined*.

As with the harbor seal solution, the emergence of dimensions to some extent resembling Extraversion, Agreeableness, Neuroticism, and Openness is consistent with previous non-human literature, as is the absence of a distinct Conscientiousness factor (Gosling & John, 1999). The Agreeable/Incompliant, Extraversion/Openness, and Dependency factors bear striking resemblance to the three previously elucidated California sea lion trait rating dimensions (Ciardelli et al., 2017), which suggests that they may have good external validity for this species in general. However, one clear potential limitation of this conclusion is that both of these studies used pinniped trainers as raters. This may therefore have contributed the good homology of factors across assessments, thus artificially inflating the apparent validity of personality dimensions. Further rating assessments of more individual animals and using raters with differing animal interactions, such as vets or researchers, would be useful to determine the extent to which these personality structures are specific to trainer-related experiences.

Cross-Species Comparisons

As predicted, pinniped trainers produced ratings which were highly reliable. Factor analyses produced several personality factors in both species, which exhibited many cross-species similarities. The sea lion Agreeable/Incompliant factor shares five trait words with the seal Stubborn/Agreeable factor, three on one pole (*temperamental, bullying,* and *incompliant*) and two on the other (*calm* and *quiet*). Other factors were not as clearly comparable, although clusters of items were shared across species. For example, items on the negative pole of sea lion Dependency and positive pole of sea lion Extraversion/Openness loaded together on the seal Interest/Neuroticism factor. Additionally, items *playful, inquisitive,* and *active,* which loaded in combination with other traits to form sea lion Extraversion/Openness, formed a separate factor in seals, Extraversion-Openness.

The absence of a distinct or combined Conscientiousness factor in either species is consistent with this factor's lack of generality across non-human species (Gosling & John, 1999). However, this is in contrast with the possible Conscientiousness-like elements present in the Routine Activity behavioral dimension found in this species (de

Vere, Lilley & Highfill, 2017), as well the finding that bottlenose dolphins can be reliably rated on Conscientiousness (Highfill & Kuczaj, 2007; Moreno, Highfill & Kuczaj, 2017). It is possible that the latter findings may be explained by the absence of factor rotation methods in these bottlenose dolphin rating studies, as a bottlenose dolphin-specific personality structure may not contain a clear Conscientiousness factor. Nevertheless, as trait rating assessments of marine mammals are still preliminary, it is plausible that current dimensional structures are not yet entirely accurate. For example, as overarching personality structures have been assessed using trait rating in only seven harbor seals (present study), 27 California sea lions (Ciardelli et al., 2017; present study), and 36 bottlenose dolphins (Highfill & Kuczaj, 2007; Moreno et al., 2017), the elucidated factors may be specific to the studied individuals. Small sample size is a clear weakness of the present study, and thus any interpretations and generalizations must be regarded as tentative. Nevertheless, the sparseness of the existing literature emphasizes the utility of additional data, even from small samples. In particular, the high similarity between California sea lion personality dimensions in the present study and those from the only previous trait rating assessment of this species (Ciardelli et al., 2017) provides some evidence of generalizability beyond the studied individuals. The present study also makes California sea lions the only marine mammal species for which personality has been assessed using trait rating and rotational statistical methods in more than one study. Overall, the presence of factors in both California sea lions and harbor seals resembling various elements and combinations of Extraversion, Neuroticism, Agreeableness, and Openness is consistent with the generality of these dimensions across the non-human personality literature (Gosling & John, 1999).

Rater Experience

A frequent assumption in trait rating assessments of non-humans is that raters who have spent more time working with the rated animals are more able to accurately rate the personality of these individuals. This is reflected in the use of criteria such as a minimum length of experience to determine 'experienced' raters in some studies (e.g. Horback, Miller & Kuczaj, 2013; Lloyd, Martin, Bornett-Gauci & Wilkinson, 2007; Lloyd, Martin, Bornett-Gauci & Wilkinson, 2008). However, evidence from assessments comparing the ability of familiar and unfamiliar raters is mixed (Feaver, Mendl & Bateson, 1986; Martau, Caine & Candland, 1985; Uher, Werner & Gosselt, 2013), while ratings have proven unreliable when raters are used who have experience with subject animals in different contexts, such as veterinarians and trainers (Highfill, Hanbury, Kristiansen, Kuczaj & Watson, 2012). Still, it remains unclear whether more extensive experience, such as that acquired by trainers and keepers, does consistently increase the reliability of personality ratings. Additionally, studies of animal personality broadly produce ratings with well above chance levels of inter-rater agreement, even when inexperienced raters are used; other measures of rater uncertainty or inaccuracy may therefore provide more nuanced information regarding the effects of rater experience.
In order to experimentally test whether this is the case, several groups of judges would need to each observe a group of animals for varied amounts of time before completing ratings. This is not unfeasible in itself, and studies in this format have been conducted, as discussed above (Martau et al., 1985; Uher et al., 2013). However, such a design becomes challenging when one wishes to give novel observers the opportunity to obtain highly varied lengths of experience (i.e. a few days versus several years) prior to rating the

subject animals. Comparing the ratings of animal caretakers with large ranges of experience with focal animals can therefore provide this type of data without requiring extensive resources.

The present study utilized existing variation in pinniped trainer experience with the study animals and in the age of these animals to examine their possible effects on rater uncertainty. There was no association between rater uncertainty and length of experience with the rated animal for both species. This is unlikely to be due to inadequate ranges of experience with the focal animals (three months to seven years) or of numbers of uncertain responses (zero to 12 per animal), both of which had substantial ranges of values. A possible explanation for this finding is that once a rater reaches a certain level of familiarity with an individual animal, they are sufficiently confident to make personality ratings. Given the previously discussed literature, this threshold may be somewhere between a few days and several months of daily experience (Martau et al., 1985; Uher et al,. 2013), which would be consistent with the minimum of three months experience trainers had with each rated animal in the present study.

Unlike the equivalent correlation in harbor seals, California sea lion age was significantly negatively correlated with rater uncertainty, indicating that raters were more confident in their ratings of older animals. This difference between these two species suggests that the effect of animal age on rating confidence may differ across taxa. During the breeding season, California sea lion females give birth to a single pup, who they stay in close proximity to while on land (Peterson & Bartholomew, 1967). Mothers typically nurse their pups for between six and twelve months, although sea lions older than one year have been observed nursing (Peterson & Bartholomew, 1967; Riedman, 1991).

Young California sea lions therefore experience an extended dependency period, during which they learn how to swim, hunt, and socialize, while remaining with their mother. This is in contrast to harbor seal pups, who are able to swim and dive almost from birth, and become independent from their mothers after nursing for only four to six weeks on average (Riedman, 1991). Pups of this species are also closer in size to adult seals than sea lion pups are to adult sea lions, and gain weight much faster to match their shorter nursing period. As a result of the shorter pre-weaning period in harbor seals, stable personality tendencies may emerge earlier in development in this species compared to taxa with a longer dependency period, such as California sea lions. If this is the case, this could explain why there was no association between harbor seal age and rater uncertainty, as all seals were weaned at the time of rating. In contrast, two of the rated California sea lions were still nursing when rated, and three further sea lions had only weaned within the last year. Juvenile sea lions may therefore have less stable or established personality traits than adults, as seems to be the case in some other species with extended sub-adult periods (Gosling & John, 1999; von Borell et al., 2016).

In some previous literature, raters have expressed their inability to rate animals because of the animal's young age, which has been equated to the rater's limited observation time (Martau, Caine & Candland, 1985). However, the results of the present study suggest that animal age and length of experience are not synonymous. Instead, rating confidence may be predominantly affected by the developmental stage of the rated animals, as animals who have reached independence may exhibit more stable behavioral tendencies. Alternatively, it is possible that this pattern reflects the extent to which raters have directly interacted with the animals, rather than solely observing them; once pups weaned from their mothers, trainers began conducting training sessions with them, and thus begin to have more concentrated, interactive exposure with these animals. As all seals in the present study had begun training sessions by the time they were rated, this may explain why there was no association between seal age and rater uncertainty. In contrast, that two sea lions had not yet begun training sessions, as well as that two young sea lions were undergoing only minimal training sessions at the time of rating, may have contributed to the negative correlation of age with rater uncertainty in this species. Most likely, the elucidated patterns, or lack thereof, are due to a combination of both of these possible explanations. Consistent with either explanation is the finding that raters reported greater overall uncertainty when rating California sea lions than harbor seals, largely due to the two un-weaned sea lion pups, who were responsible for 45% of the total uncertain responses in this species.

Other facilities may have different setups for their pinniped populations and trainers, such that it may relatively easy to tease apart the relative effects of trainer interaction and personality trait stability. For example, some facilities have dedicated off-exhibit space for mother-pup pairs which facilitate observations of nursing pups by trainers prior to conducting training sessions with them. If the same negative correlation between rater uncertainty and animal age emerged in this setting, it would support the developmental explanation. If the correlation was not significant, it would suggest that the finding in the present study is due to a lack of trainer-pup interaction prior to pup weaning.

Dominance & Age

Both California sea lions and harbor seals are known to form dominance hierarchies, broadly based on sex and age (Riedman, 1991). California sea lions are highly sexually dimorphic, and the most dominant males maintain territories during the breeding season (Riedman, 1991), while harbor seals display minimal sexual dimorphism, but do still exhibit sex- and age-based hierarchies (Sullivan, 1982). Thus, adult males are most dominant in both species, with young pups least dominant, regardless of sex. Consistent with this literature, dominance and age were highly positively correlated for both species in the present study. Although some association between age and dominance ranking was expected, these effects were extremely large. In particular, sea lion ages and dominance rankings were almost perfectly correlated.

However, this strong correlation raises difficulties when interpreting their correlation with personality dimensions. Age is thought to be directly related to a wild individual's position in a dominance hierarchy due to the general increase in body size and weight, and therefore ability to win agonistic interactions (Godsell, 1991; Neumann, 1999). This relationship makes it difficult to determine the extent to which personality is causally related to dominance and/or age. For example, an animal becoming larger with increased age may facilitate a higher position in a dominance hierarchy; in turn, this may influence their personality tendencies. Alternatively, increased age could contribute to changes in personality, which may facilitate movement up the dominance hierarchy. In the latter scenario, increased body size with age may have little effect on dominance, or could add to the positive effects of personality change by also contributing to more successful agonistic interactions.

Presumably, at some point an increase in age would no longer be associated with concurrent increase in dominance ranking for two reasons: one, because there is a ceiling effect (i.e. once an animal reaches the top of the dominance hierarchy it is not possible to obtain a higher ranking, females cannot supersede the most dominant males, and there is an upper limit on possible body size), and two, because of physical deterioration in older age. However, most wild pinnipeds may be unlikely to live to a sufficiently old age to experience such deterioration in physical condition that their dominance ranking decreases as a result, and all of the individual animals in the present study are juveniles or are still reproductively active. Studying the personality of post-reproductive animals may therefore provide a way to distinguish between some of the possible causal mechanisms discussed above. For example, if certain personality traits are more important than body condition, then an individual with those traits would be expected to retain their dominance ranking for a period of time after experiencing some loss of body condition. If body size/condition is more influential, then such an individual would be expected to move down in the hierarchy immediately in relation to change in size/condition. Conducting longitudinal studies of both personality and dominance will likely be most effective for determining the direction of the potential causal relationship between these two variables. There may also be interactions between these factors, such as if one is more important for increasing one's rank, while the other is more influential in maintaining that rank, that were not possible to detect in the present study.

Even in the absence of this type of data, a number of interesting correlations emerged between several personality factors and both dominance and age. For both species, age and dominance were both negatively correlated with the Boldness behavioral factor. This indicates that as age and dominance ranking increased, the frequency with which animals moved around on land, and engaged in tactile, scanning, and aggressive behaviors decreased (de Vere et al., 2017). Both species' Boldness factors were interpreted as containing some Extraversion-like tendencies; the lower Boldness scores in older pinnipeds is therefore consistent with findings in cats, snow leopards (Gartner et al., 2014), and another population of California sea lions (Ciardelli et al., 2017), in which older animals had lower scores on personality dimensions resembling Extraversion.

Both dominance and age were positively correlated with sea lion Dependency; more dominant, older sea lions were rated as more determined and less dependent, fearful, and nervous. This makes intuitive sense, as neither species maintains long-term, stable social groupings year round, and thus animals must be able to make decisions independently or risk not surviving. Less fearful sea lions may also be more likely to engage in confrontations with individuals of similar or higher dominance rankings, and therefore be more likely to seek out opportunities to advance in the dominance hierarchy. The finding that both Dependency and Boldness were each correlated with dominance is also consistent with the significant negative correlation between these personality dimensions in this species.

Finally, seal Extraversion-Openness was negatively correlated with dominance, but not age, with more dominant seals rated as less playful, inquisitive, and active, and more predictable. This suggests that pinniped trainers perceived animals who were interested in their environment and behaved unpredictably as lower in the dominance hierarchy. This appears consistent with the idea that once animals have established themselves at the top of a hierarchy, they no longer necessarily need to exhibit behavioral indicators often associated with dominance, such as high aggressiveness and reactivity. This may particularly be the case for captive populations, in which social groupings tend to be smaller and less variable than those in the wild, and may therefore have more stable dominance hierarchies. However, the absence of a significant correlation with age in harbor seals is in contrast to existing non-human literature suggesting that in many species, Extraversion and/or Openness scores decrease with age (Ciardelli et al., 2017; Eckardt et al., 2015; Gartner et al., 2014; King, Weiss & Sisco, 2008).

One personality dimension produced by both behavioral coding and trait rating methods was therefore correlated with dominance in both species. This is significant, as these associations emerged despite the intentional exclusion of *dominant* and *submissive* trait words. However, all of the findings discussed here are limited by both the relatively small number of total animals assessed, and the somewhat limited variability in age and dominance scores. For instance, individuals of both species were broadly either young juveniles or older adults, with very few sub- or young adults, particularly for harbor seals (Figures 3a, 4a, 5a, 6a). Additionally, harbor seal dominance rankings had a very small range, as no individuals were rated as ‘high’ dominance, so the elucidated patterns may not be reflective of those spanning a more varied hierarchy.

Convergent Validity

The present study produced somewhat mixed evidence for the cross-method convergent validity of personality dimensions in two marine mammal species, with California sea lions exhibiting better convergence. In this species, the trait rating Dependency factor was negatively correlated with the behavioral Boldness factor. This indicates that raters judged sea lions who engaged in more open mouth, social play, tactile, and aggressive behaviors (de Vere et al., 2017) as being more *quiet*, *fearful*, *dependent* and *nervous*. This is a significant finding, as it casts doubt on the original interpretation of this behavioral factor. The combination of interactive behaviors (e.g. tactile, aggression, social play) and movement on land and in and out of the water was interpreted as indicative of confidence and boldness, and to some extent curiosity (de Vere et al., 2017). However, in light of the correlation of scores on this factor with scores on the Dependency factor, it seems that it may be more accurate to interpret these behavioral tendencies as indicative of impulsive and Neuroticism-like traits. For example, in this population, the movement in/out of the water variable included in factor analyses reflected the time, in seconds, an animal spent moving in and out of the water. It therefore does not simply reflect the extent to which an animal hauled out on land; it actually reflects the duration of time the individual spent moving on and off land. An individual who might be considered confident or bold would be expected to haul out onto land and remain there throughout potential disturbances, and thus would have spent a very small amount of time actually engaging in movement in/out behaviors. This type of individual would have lower scores on this factor as a result. Similarly, interactive behaviors, such as tactile and aggressive interactions, may reflect the need to seek out comfort via social connections and instability, respectively, rather than confidence. These new interpretations are more in line with Neuroticism-like tendencies, such as excitableness and instability (Goldberg, 1990). While harbor seal Boldness was not substantially correlated with any of the seal rating factors, the extremely similar combinations of behaviors on both seal and sea lion Boldness (de Vere et al., 2017) suggest that seal Boldness should also be re-interpreted. The use of multiple methods has therefore provided a more objective way to distinguish between several potential explanations.

The positive correlation between the Extraversion/Openness rating factor and Routine Activity behavioral factor was also approaching significance. Sea lions rated as more *intelligent*, *curious*, and *active* were also more alert and exhibited more pattern swimming and less resting (de Vere et al., 2017). This makes some theoretical sense, based on the *activity* and *alert* element of the Extraversion/Openness factor. However, this correlation also suggests that the Routine Activity factor may be more reflective of Extraversion and Openness-like tendencies than was initially thought. In particular, it was difficult to interpret the negative pole of Routine Activity in the absence of other sources of information (de Vere et al., 2017), but in light of the correlation of resting and maintenance behaviors with *unoriginal*, *lazy*, and *solitary* traits, it suggests that animals low on Routine Activity exhibit low Openness-like tendencies (Goldberg, 1990).

Unlike the sea lion results, there were no significant correlations between any harbor seal behavioral coding and trait rating personality factors. This cannot simply be explained by a complete lack of rating validity for this species, as several of the individual trait-behavior correlations were significant, and were extremely similar to the equivalent sea lion correlations. This suggests that trainers did base their ratings on harbor seal behavior, but that the way in which traits and behaviors clustered into factors in this sample did not correspond well across methods. The good cross-method correspondence of sea lion personality dimensions lends further support to the argument that something about the harbor seal methodology contributed to a lack of general factor convergence. It is possible that the small sample size (25 total ratings, by three or four raters of seven seals) was not sufficient to detect any patterns; if a larger sample size were tested, perhaps moderate correlations would become significant, such as the positive correlations of Boldness (now Impulsivity) and both Interest/Neuroticism and Dependency (Table 12). Alternatively, perhaps the extent of rater experience with each species as a whole affects rater ability to accurately detect underlying dimensions; on average, trainers had six years of experience with California sea lions (standard deviation 2.92 years), but only 3.9 years with harbor seals (standard deviation 4.34 years). This being said, raters had at least 6 months of experience with harbor seals, and prior studies have included raters with extensive daily experience over this period (e.g. Lloyd et al., 2007). Another possible explanation is that the behavior of harbor seals in the present study during interactions with their trainers is not representative of their overall behavioral tendencies or underlying personality structure. For example, as harbor seals are known to be less gregarious and social than many pinniped species, including sea lions (Riedman, 1991), perhaps their interactions with trainers reflect only their social tendencies, and not other aspects of their personality. Trait ratings made by trainers may therefore still be based on observed behaviors, thus explaining the correlation of individual traits and behaviors, but their dimensional structure could reflect only training-related personality factors. In contrast, sea lion behavior during training sessions may be representative of their general personality, thus producing the good correspondence between their rating and behavioral factors. It is not possible to test this hypothesis with the current data, so future research could incorporate behavioral assessments of seals during training sessions, to determine whether any personality factors produced from such observations correspond better with trainer ratings. Additionally, if this explanation is valid, one would expect ratings made by people in non-trainer roles to show greater cross-method convergence; incorporating raters with other sources of experience might therefore prove useful.

The traits used in the present study's personality questionnaires were not defined using species-specific behaviors. Some have advocated for the use of such behavioral adjectives rather than trait adjectives, due to evidence that the latter have been associated with worse cross-method convergence in some non-human literature (e.g. Uher & Asendorpf, 2008). It has been argued that the use of behavioral adjectives results in greater external validity and decreases subjectivity, as raters should theoretically base their ratings of a trait on the behaviors described in that trait's definition, rather than their own interpretation of a typical dictionary definition (Uher & Asendorpf, 2008). However, all trait rating inherently involves some subjectivity, regardless of whether adjective or behavioral traits are used; when the former are used, raters make their own interpretations regarding which elements of their experience with an animal reflect each trait, while the creators of the questionnaire decide which behaviors are indicators of each trait when the latter are used. Both therefore involve inherently subjective interpretations of the behavioral indicators of each trait, but one would expect those interpretations made by 'experts' to be more accurate, and therefore produce more reliable and/or accurate ratings. However, few studies have actually compared the relative validity of ratings produced from adjective and behavioral traits; most simply use one or the other, which precludes comparisons between the two. Nevertheless, there are many examples of both behavioral and adjective trait ratings showing good convergence with behavioral measures (Barnard et al., 2016; Lloyd et al., 2007; Pederson et al., 2005), even within studies which do compare them (Uher & Asendorpf, 2008). It could also be argued that ratings of trait adjectives which show good convergence with behavioral measures, as was the case in the present study, may provide stronger evidence of convergent validity; such convergence demonstrates that raters can make 'subjective' interpretations of trait indicators in highly similar and broadly accurate ways, which would be unlikely if based solely on individual raters' anthropomorphic interpretations of traits.

With the exception of the broad harbor seal dimensions, there was therefore substantial cross-method convergence of personality in the current study. This is consistent with much existing non-human literature examining the validity of 'subjective' ratings, although little of this research has been conducted in marine mammals. Perhaps most relevant to the present study is the finding that both bottlenose and rough-toothed dolphins with higher scores on a Curiosity rating factor looked longer at an unpredictable visual stimulus (Lilley et al., 2018). Behavioral observations of many primate species have shown good correspondence with trait ratings; for example, chimpanzee Agonistic behaviors were positively correlated with Dominance and negatively with Agreeableness rating factors (Pederson et al., 2005). Cross-method convergence of broad personality dimensions has also emerged in domestic species, such as horses (Lloyd et al., 2007) and dogs (Barnard et al., 2016). The current study's findings therefore further support the use of trait rating as a valid method for assessing non-human animal personality, as well as contributing specific support for its use in marine mammals.

Conclusions

The present study provides the first evidence for the cross-method validity of personality dimensions in any pinniped species. In particular, there is strong crossmethodological support for a Neuroticism-like and a combined Extraversion/Openness factor in California sea lions. The convergence of sea lion behavioral coding and trait rating factors hints at an interesting possibility: that trainer-animal interactions may be representative of the personality of some species but not others. The behavioral Boldness factor was relabeled (from de Vere et al., 2017) as Impulsivity, as its cross-method convergence with sea lion Dependency suggested that behaviors on this factor might be more indicative of instability than confidence. This is a clear demonstration of the subjectivity involved in the behavioral coding method, despite this method's traditional reputation as highly objective, and the usefulness of multiple methods for guiding the interpretation of non-human personality dimensions.

Human raters may be capable of rating the emotional repertoires experienced by non-human species. The present study was an extremely preliminary assessment of this possibility, but the high level of agreement of several raters on over twenty emotional states in both species suggests that the rating method holds some promise in this area. Much as existing studies have applied the 'expert' nomination method for identifying species-relevant personality traits, people who have extensive experience with and/or exposure to a species may be best placed to identify the emotional states that they are most likely to experience. These species-specific emotional repertoires could then be tested for validity using personality-based behavioral measures and available biological indicators of emotional experiences.

A clear, overarching limitation of the present study is the small number of raters surveyed. It is much more difficult to recruit large numbers of raters for wild species than it is for domestic animals, with which many people have substantial experience as pet owners, because of the inherent difficulty involved in repeatedly observing individually identifiable wild animals. A system or network through which people with extensive experience with wild animals could be recruited, by the species of interest, would be extremely useful for expanding this type of research in the future. Nevertheless, this study provides novel information to several branches of non-human research, including emotions, personality, and the utility of cross-method approaches. In particular, adding to the marine mammal personality literature will hopefully stimulate further research to develop and validate taxa-specific personality assessments, as has been achieved for several other taxa (e.g. Weiss et al., 2007; Wiener & Haskell, 2016). Such tools could then be used to examine whether associations between personality and other variables, which have been elucidated in other species, exist in marine mammals; these factors include important welfare-related outcomes, such as interactions with environmental stimuli (Lilley et al., 2018) and engagement in stereotypic behaviors (Gottlieb, Capitanio & McCowan, 2013). Given that marine mammals broadly have rich social lives, possess complex cognitive abilities, but also exhibit great inter-species variation in life history and ecological characteristics, this taxon has the potential to provide unique insights into the development, plasticity, and evolution of personality.

APPENDIX A – Tables & Figures

Table A1. *Subject animal demographic information at the time of data collection.*

Animal	**Age (years)**	**Sex**	**Species**
1	11	F	Seal
2	10	F	Seal
3	2	F	Seal
4	1	F	Seal
5	0.6	F	Seal
6	13	M	Seal
7	0.6	M	Seal
8	20	F	Sea lion
9	9	F	Sea lion
10	4	F	Sea lion
11	2	F	Sea lion
12	1	F	Sea lion
13	20	M	Sea lion
14	20	M	Sea lion
15	13	M	Sea lion
16	2	M	Sea lion
17	0.5	M	Sea lion
18	0.5	M	Sea lion

Table A2. *Years of experience of raters with harbor seals and California sea lions. (* denotes rater not included in harbor seal analyses).*

Rater	**Years of experience**	
	Harbor seals	**California sea lions**
1	1	5
2	0.17*	1.5
3	1	6
4	7	7
5	10	10

Table A3. *Mean ratings, standard deviations, and rater percentage agreement for rated emotional states in harbor seals and California sea lions, organized by source dimension. (Bold = agreement <1 standard deviation, * = mean rating equal to or less than 3)*

State word origin	**State word**	**Harbor seals**			**California sea lions**		
		Mean	Standard deviation	Percentage agreement	Mean	Standard deviation	Percentage agreement
PANAS	Enthusiastic	**4**	0.82	85.7	**5.8**	0.83	85.7
	Interested	**6.25**	0.5	92.9	5.8	1.64	74.3
	Determined	5	1.41	73.8	**4.6**	0.89	85.7
	Excitable	5.25	1.26	78.6	**6.2**	0.84	85.7
	Inspired	**3***	1	81	**2.75***	0.96	83.3

Error! No text of specified style in document. (continued).

PANAS	Alert	**6.5**	0.58	90.5	**6.8**	0.45	94.3
	Active	**6**	0.82	85.7	**6.6**	0.55	91.4
	Proud	3	1.83	66.7	2.6	1.52	74.3
	Energetic	5.25	1.26	78.6	**6.4**	0.89	85.7
	Confident	5.25	1.71	69.1	5.25	1.5	73.8
	Fearful	**6.5**	0.58	90.5	**5.6**	0.89	85.7
	Afraid	**6.5**	0.58	90.5	5.4	1.14	80
	Upset	3	2.83	52.4	3.8	1.64	71.4
	Distressed	5.5	1.29	76.2	4.2	1.30	77.1
	Jittery	**5.5**	1	85.7	4.25	1.5	78.6
	Nervous	**6.75**	0.5	92.9	**5.2**	0.84	85.7
	Ashamed	**1.5***	0.58	90.5	**1.4***	0.55	91.4
	Guilty	**2***	0.82	85.7	2*	1.23	80
	Irritable	5.5	1.29	76.2	**5.4**	0.55	91.4
	Hostile	3.33	3.22	42.9	4	1.87	68.8
Openness	Curious	**6**	0.82	85.7	**6.6**	0.55	91.4
Conscientiousness	Cautious	**6.75**	0.5	92.9	5.2	1.30	77.1
	Lazy	**6**	0.82	85.7	**5.4**	0.55	91.4
	Reckless	3	2.65	52.4	2.67	2.08	61.9
Extraversion	Playful	**5.25**	0.5	92.9	**5.8**	0.84	85.7
	Quiet	**7**	0	100	**4.5**	0.57	90.5

Error! No text of specified style in document. (continued).

Extraversion	Depressed	2.75	1.71	69.1	3.2	1.79	68.6
Agreeableness	Sensitive	**6.33**	0.58	90.5	2.33*	1.55	81
	Protective	4.5	1.73	71.4	**6.2**	0.45	94.3
	Gentle	**6.5**	0.58	90.5	4.8	1.10	82.9
	Greedy	3.33	2.08	61.9	3.75	2.63	54.8
	Friendly	4	2	71.4	3.8	2.28	60
	Cooperative	6	1.16	81	**6.2**	0.84	85.7
	Stubborn	**6**	0.82	85.7	5.6	1.14	80
	Selfish	3.75	3.21	45.2	4.5	2.65	52.4
	Sympathetic	**1.5***	0.58	90.5	1.8*	1.30	80
	Aggressive	3.5	1.29	76.2	**5.4**	0.89	88.6
	Tolerant	**5.75**	0.96	83.3	6	1.23	80
	Defiant	4.5	1.29	76.2	4.25	2.5	54.8
Neuroticism	Jealous	**2***	1	81	**2.75***	0.96	83.3
	Anxious	**6**	0.82	85.7	**5.4**	0.89	88.6
	Timid	5.5	1.73	71.4	4.8	1.48	74.3
	Unexcitable	**5**	1	81	4.6	2.07	62.9
	Calm	**6**	0.82	85.7	**6.25**	0.96	83.3
	Unemotional	2.5	1.73	71.4	3	2.16	61.9
	Vulnerable	4.75	2.06	64.3	3.6	1.52	77.1

Table A4. *Initial Principal Component Analysis (PCA) and Regularized Exploratory Factor Analysis (REFA) results for harbor seal trait ratings. (Bold denotes factor loading greater than 0.35, * denotes highest loading across all three factors).*

Trait Item	**PCA**			**REFA**		
	Factor 1	**Factor 2**	**Factor 3**	**Factor 1**	**Factor 2**	**Factor 3**
Curious	**-0.787***	0.094	**-0.391**	**-0.7222***	**0.3577**	0.0483
Intelligent	-0.070	0.144	**-0.739***	0.0748	**0.8003***	-0.2186
Inquisitive	**-0.672***	0.115	**-0.566**	**-0.5572**	**0.5840***	-0.0509
Simple	0.052	-0.330	**0.391***	0.0153	-0.3446*	-0.1420
Unoriginal	0.033	-0.152	**0.381***	**0.4848**	**0.5701***	0.0756
Alert	**0.426**	**0.392**	**-0.537***	**0.9232***	0.0531	-0.0321
Cautious	**0.898***	0.111	0.100	-0.0500	**-0.5076***	**0.4255**
Lazy	0.032	0.218	**0.557***	0.0310	-0.2718*	0.2059
Predictable	0.081	0.093	**0.423***	-0.1320	**0.5012***	0.1502
Active	-0.195	0.303	**-0.491***	-0.2898	**0.6999***	0.0604
Playful	**-0.401**	0.283	**-0.681***	0.1334	**-0.7824***	-0.1115
Quiet	-0.121	**-0.516***	0.173	-0.0691	-0.0746	**-0.4386***
Solitary	-0.337	-0.103	0.210	-0.3499*	-0.2468	0.0128
Sensitive	**0.393**	**0.529***	-0.139	**0.3834***	0.1945	**0.3563**
Gentle	0.183	**-0.734***	0.212	0.2138	-0.2025	**-0.6728***

Error! No text of specified style in document. (continued).

Stubborn	0.053	**0.701***	0.268	-0.0101	-0.0854	**0.7415***
Incompliant	**0.459**	**0.377**	**0.552***	**0.3696**	**-0.3569**	**0.4836***
Demanding	**-0.594***	0.277	-0.002	**-0.6274***	-0.0305	**0.4057**
Temperamental	**0.496**	**0.670***	-0.136	**0.4793**	0.2188	**0.5332***
Tolerant	**-0.405**	**-0.677***	-0.171	-0.3232	0.1405	**-0.6796***
Calm	-0.102	**-0.791***	0.318	-0.1025	**-0.3644**	**-0.6236***
Stable	**-0.697***	-0.266	-0.134	**-0.6729***	-0.0382	-0.1983
Fearful	**0.900***	0.101	0.078	**0.9276***	0.0646	-0.0450
Bullying	-0.337	**0.649***	-0.055	**-0.3668**	0.0376	**0.6439***
Dependent	**0.618***	0.190	-0.224	**-0.6549***	0.3364	-0.0310
Enthusiastic	**-0.624***	0.319	-0.345	**-0.5686***	**0.3543**	0.2610
Interested	**-0.809***	-0.045	**-0.394**	**-0.7259***	**0.3990**	-0.1143
Jittery	**0.873***	0.080	0.166	**0.8597***	-0.1004	0.0027
Nervous	**0.937***	0.158	0.171	**0.9365***	-0.0579	0.0487
% variance	28.934	16.511	9.439	7.7068	4.0899	4.2442
Alpha	0.936	0.827	0.712	0.898	0.8525	0.817

Table A5. *Final Regularized Exploratory Factor Analysis Structure for harbor seals, with items: lazy, simple, unoriginal, and solitary removed. (Bold denotes factor loading greater than 0.35, * denotes highest loading across all three factors)*

Trait Item	**Interest/ Neuroticism**	**Stubborn/ Agreeable**	**Extraversion/ Openness**
Interested	**0.7109***	-0.1176	**0.404**
Curious	**0.7014***	0.0326	**0.4017**
Demanding	**0.6357***	**0.3751**	-0.0414
Stable	**0.6265***	-0.2684	0.0665
Enthusiastic	**0.5525***	0.2639	**0.3737**
Nervous	**-0.9449***	0.0451	-0.014
Fearful	**-0.9400***	-0.0382	0.0931
Cautious	**-0.9265***	-0.0151	0.0476
Jittery	**-0.8725***	-0.0264	-0.0572
Dependent	**-0.6610***	-0.0123	**0.3637**
Sensitive	**-0.3864***	**0.3788**	0.2221
Stubborn	0.0484	**0.7498***	-0.1616
Bullying	**0.3798**	**0.6170***	0.0308
Temperamental	**-0.455**	**0.5797***	0.1832
Incompliant	-0.3466	**0.4595***	**-0.3873**
Gentle	-0.23	**-0.6847***	-0.2173
Tolerant	0.3186	**-0.6492***	0.1024

Error! No text of specified style in document. (continued).

Calm	0.0937	**-0.6421***	**-0.3564**
Quiet	0.0134	**-0.5155***	0.0411
Playful	0.2515	0.0444	**0.7814***
Intelligent	-0.0524	-0.1069	**0.6957***
Inquisitive	**0.5491**	-0.0229	**0.5778***
Active	0.1296	0.1748	**0.5327***
Alert	**-0.4516**	0.1898	**0.4860***
Predictable	0.0467	0.2764	**-0.4636***
Cronbach's alpha	0.930	0.833	0.711
% Variance	7.8005	3.9977	3.8177

Table A6. *Spearman's rank order correlations between final harbor seal trait rating factors.*

	Interest/ Neuroticism	Stubborn/ Agreeable	Extraversion-Openness
Interest/Neuroticism	-	0.1454	0.1624
Stubborn/Agreeable	-	-	0.1492
Extraversion-Openness	-	-	-

Table A7. *Initial Principal Component Analysis (PCA) and Regularized Exploratory Factor Analysis (REFA) results for California sea lion trait ratings. (Bold denotes factor loading greater than 0.35, * denotes highest loading across all three factors)*

Trait Items	**PCA**			**REFA**		
	Factor 1	**Factor 2**	**Factor 3**	**Factor 1**	**Factor 2**	**Factor 3**
Curious	**0.586***	0.126	0.212	0.0387	**0.5640***	-0.0490
Intelligent	0.331	-0.288	0.059	-0.0340	0.3394*	0.2216
Inquisitive	**0.785***	0.332	0.189	-0.0328	**0.7490***	-0.2791
Simple	**-0.437**	**0.499***	-0.281	-0.1519	**-0.4046**	**-0.4379***
Unoriginal	**-0.771***	-0.011	-0.080	0.1502	**-0.7875***	0.0305
Alert	**0.573***	-0.023	**0.521**	**0.4032***	**0.3872**	0.0378
Aggressive	0.019	**-0.366**	**0.707***	**0.7499***	-0.2128	**0.4031**
Lazy	**-0.692***	-0.153	-0.308	-0.1299	**-0.5955***	0.1375
Predictable	-0.302	-0.154	**-0.683***	**-0.6678***	-0.0246	0.0993
Active	**0.783***	0.003	**0.599**	**0.3954**	**0.6342***	0.0620
Playful	**0.627***	**0.525**	0.301	0.1560	**0.4991***	**-0.4464**
Energetic	**0.827***	0.054	**0.653**	**0.4533**	**0.6503***	0.0106
Quiet	**-0.462**	**0.362**	**-0.488***	-0.3482	**-0.3839**	**-0.3940***
Solitary	**-0.405***	-0.322	0.133	0.2508	**-0.4531***	0.2739
Cooperative	-0.247	-0.247	**-0.728***	**-0.7578***	0.0940	0.1868
Protective	-0.301	**-0.469***	-0.238	-0.1755	-0.1675	**0.3952***

Error! No text of specified style in document. (continued).

Incompliant	-0.079	0.104	**0.817***	**0.9606***	**-0.4693**	-0.0445
Demanding	0.334	**-0.398**	**0.576***	**0.4824***	0.2196	**0.4437**
Temperamental	**0.582**	**0.369**	**0.625***	**0.5360***	0.3140	-0.3167
Excitable	**0.849***	-0.056	**0.391**	0.1422	**0.8015***	0.0924
Anxious	**0.570**	**0.447**	**0.628***	**0.5642***	0.2686	**-0.4208**
Calm	**-0.535**	0.107	**-0.804***	**-0.7005***	-0.2854	-0.1699
Stable	**-0.555**	-0.109	**-0.784***	**-0.7145***	-0.2447	0.0817
Fearful	**0.570**	**0.668***	**0.398**	0.3179	0.3256	**-0.6623***
Bullying	0.276	**-0.477**	**0.764***	**0.7239***	0.0763	**0.5389**
Dependent	0.137	**0.856***	-0.015	0.0004	0.0069	**-0.8230***
Enthusiastic	**0.850***	0.139	0.132	-0.1260	**0.8681***	-0.1158
Determined	-0.083	**-0.501***	0.029	0.0370	-0.0325	**0.4529***
Irritable	0.141	-0.003	**0.647***	**0.6415***	-0.0982	0.0552
Nervous	0.334	**0.647***	**0.574**	**0.6018**	-0.0245	**-0.6391***
Cronbach's alpha	0.901	0.692	0.907	0.923	0.917	0.751
% variance	35.820	13.815	10.224	6.5531	5.8954	4.1411

Table A8. *Final Regularized Exploratory Factor Analysis Structure for California sea lions, with items: intelligent, simple, and protective removed. (Bold denotes factor loading greater than 0.35, * denotes highest loading across all three factors)*

Trait Items	**Agreeable/ Incompliant**	**Extraversion/ Openness**	**Dependency**
Cooperative	**0.7577***	0.0928	0.1894
Stable	**0.7176***	-0.2403	0.0933
Calm	**0.7003***	-0.2842	0.1817
Predictable	**0.6724***	-0.0249	0.095
Quiet	**0.4388***	**-0.3913**	**-0.4249**
Incompliant	**-0.9566***	**-0.4634**	-0.0476
Aggressive	**-0.7507***	-0.2178	**0.3918**
Bullying	**-0.7295***	0.0679	**0.5473**
Irritable	**-0.6323***	-0.0882	0.0713
Anxious	**-0.5665***	0.2702	**-0.4186**
Temperamental	**-0.5318***	0.3278	-0.2909
Demanding	**-0.4805***	0.2221	**0.4739**
Alert	**-0.4147***	**0.3591**	0.0098
Enthusiastic	0.1091	**0.8538***	-0.1077
Excitable	-0.1555	**0.7905***	0.0986
Inquisitive	0.0277	**0.7573***	-0.2491
Energetic	**-0.4609**	**0.6451***	0.0211

Error! No text of specified style in document. (continued).

Active	**-0.3962**	**0.6435***	0.0893
Curious	-0.0384	**0.5813***	0.0016
Playful	-0.1468	**0.5335***	**-0.4045**
Unoriginal	-0.138	**-0.7678***	0.0228
Lazy	0.135	**-0.5905***	0.15
Solitary	-0.2522	**-0.4817***	0.2431
Determined	-0.0504	-0.053	**0.3997***
Dependent	0.0112	0.0329	**-0.8135***
Fearful	-0.3163	0.3296	**-0.6661***
Nervous	**-0.5981**	-0.0244	**-0.6592***
Cronbach's alpha	0.907	0.906	0.705
% variance	6.5120	6.6094	4.7190

Table A9. *Spearman's rank order correlations between final California sea lion trait rating factors.*

	Agreeable/ Incompliant	Extraversion/ Openness	Dependency
Agreeable/Incompliant	-	0.0568	-0.1050
Extraversion/Openness	-	-	-0.3981
Dependency	-	-	-

Table A10. *Spearman's rank order correlations of harbor seal trait rating and behavioral coding factors with animal age and rated dominance ranking. (Underlined denotes p<0.1, bolded denotes p<0.05, * denotes p<0.001)*

		Age	Dominance Ranking
Behavioral Coding Factors	Boldness	**-0.955***	<u>-0.685</u>
	Routine Activity	0.631	0.252
Trait Rating Factors	Interest/Neuroticism	-0.360	-0.090
	Stubborn/Agreeable	-0.252	-0.054
	Extraversion-Openness	-0.577	<u>-0.721</u>

Table A11. *Spearman's rank order correlations of California sea lion trait rating and behavioral coding factors with animal age and rated dominance ranking. (Underlined denotes p<0.1, bolded denotes p<0.05, * denotes p<0.001)*

		Age	Dominance Ranking
Behavioral Coding Factors	Boldness	**-0.793***	**-0.851***
	Routine Activity	0.088	0.047
Trait Rating Factors	Agreeable/Incompliant	-0.267	-0.220
	Extraversion/Openness	-0.277	-0.257
	Dependency	**0.825***	**0.799***

Table A12. *Spearman's rank order correlations between harbor seal behavioral coding and trait rating factors. No correlations p<0.1.*

		Trait Rating Factors		
		Interest/ Neuroticism	Stubborn/ Agreeable	Extraversion-Openness
Behavioral Coding Factors	Boldness	0.429	0.143	0.464
	Routine Activity	-0.107	-0.250	0.071

Table A13. *Spearman's rank order correlations between California sea lion behavioral coding and trait rating factors. (Underlined denotes p<0.1, bolded denotes p<0.05, * denotes p<0.001)*

		Trait Rating Factors		
		Agreeable/ Incompliant	Extraversion/ Openness	Dependency
Behavioral Coding Factors	Boldness	0.209	0.209	**-0.727**
	Routine Activity	0.455	<u>0.582</u>	-0.009

Table A14. *Spearman's rank order correlations between five harbor seal behaviors and four rated traits. Shaded cells denote a priori predicted behavior-trait correlations. (Underlined denotes p<0.1, bolded denotes p<0.05, * denotes p<0.001)*

	Rated Traits				
Behaviors		Playful	Curious	Active	Alert
	Alone play	**0.883**	0.536	**0.757**	0.321
	Social play	0.252	0.357	0.09	0.536
	Tactile	0.414	0.643	0.559	0.321
	Resting	<u>-0.685</u>	0.294	<u>-0.739</u>	-0.429
	Scanning	0.667	**0.821**	0.595	0.143

Table A15. *Spearman's rank order correlations between five California sea lion behaviors and four rated traits. Shaded cells denote a priori predicted behavior-trait correlations. (Underlined denotes p<0.1, bolded denotes p<0.05, * denotes p<0.001)*

	Rated Traits				
Behaviors		Playful	Curious	Active	Alert
	Alone play	<u>0.0528</u>	0.315	0.446	0.320
	Social play	0.445	0.114	0.437	0.183
	Tactile	**0.620**	0.498	0.073	-0.087
	Resting	-0.500	-0.260	**-0.624**	-0.562
	Scanning	**0.611**	0.429	0.469	0.482

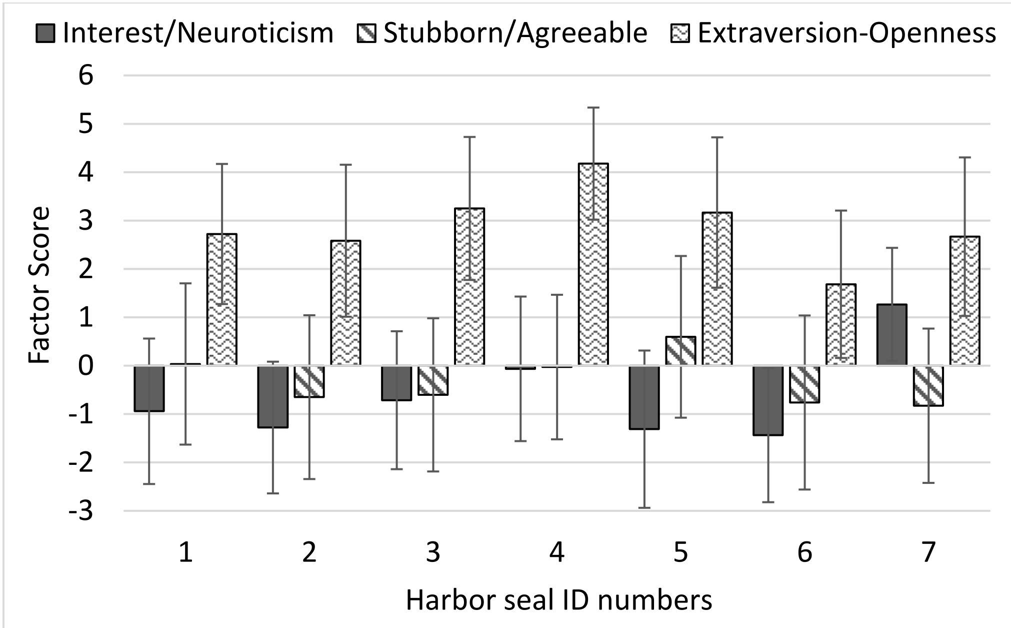

Figure A1. Harbor seal scores on final trait rating factors, with standard error bars.

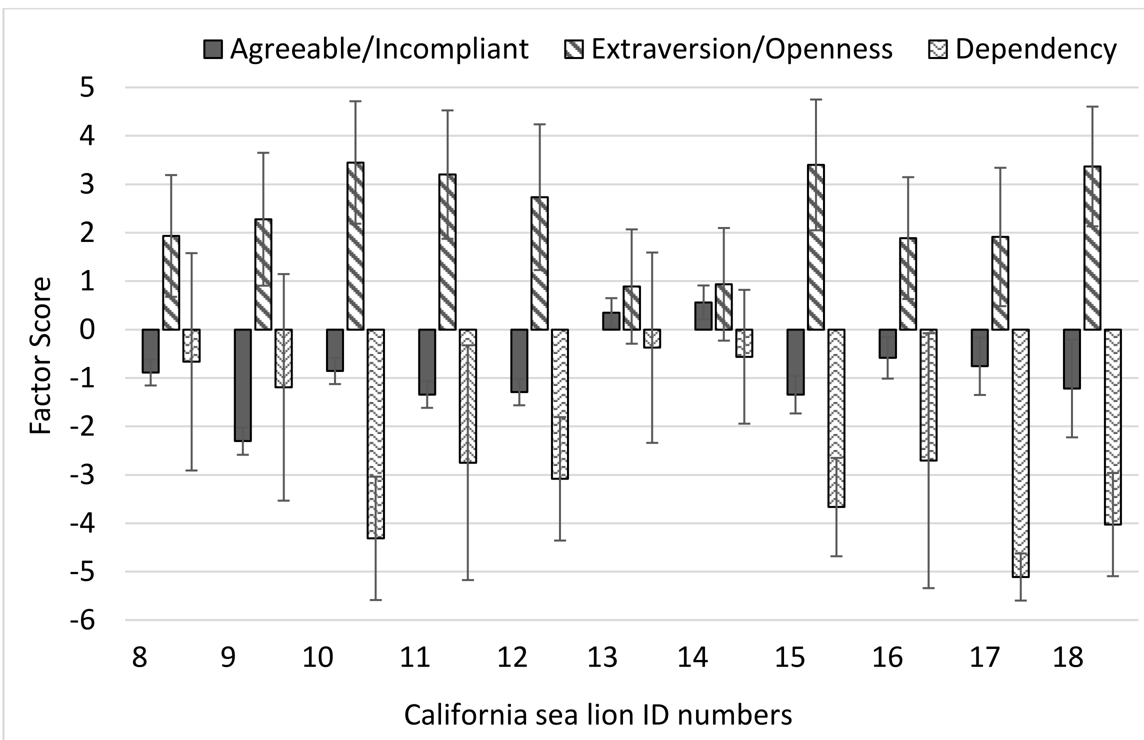

Figure A2. California sea lion scores on final trait rating factors, with standard error bars.

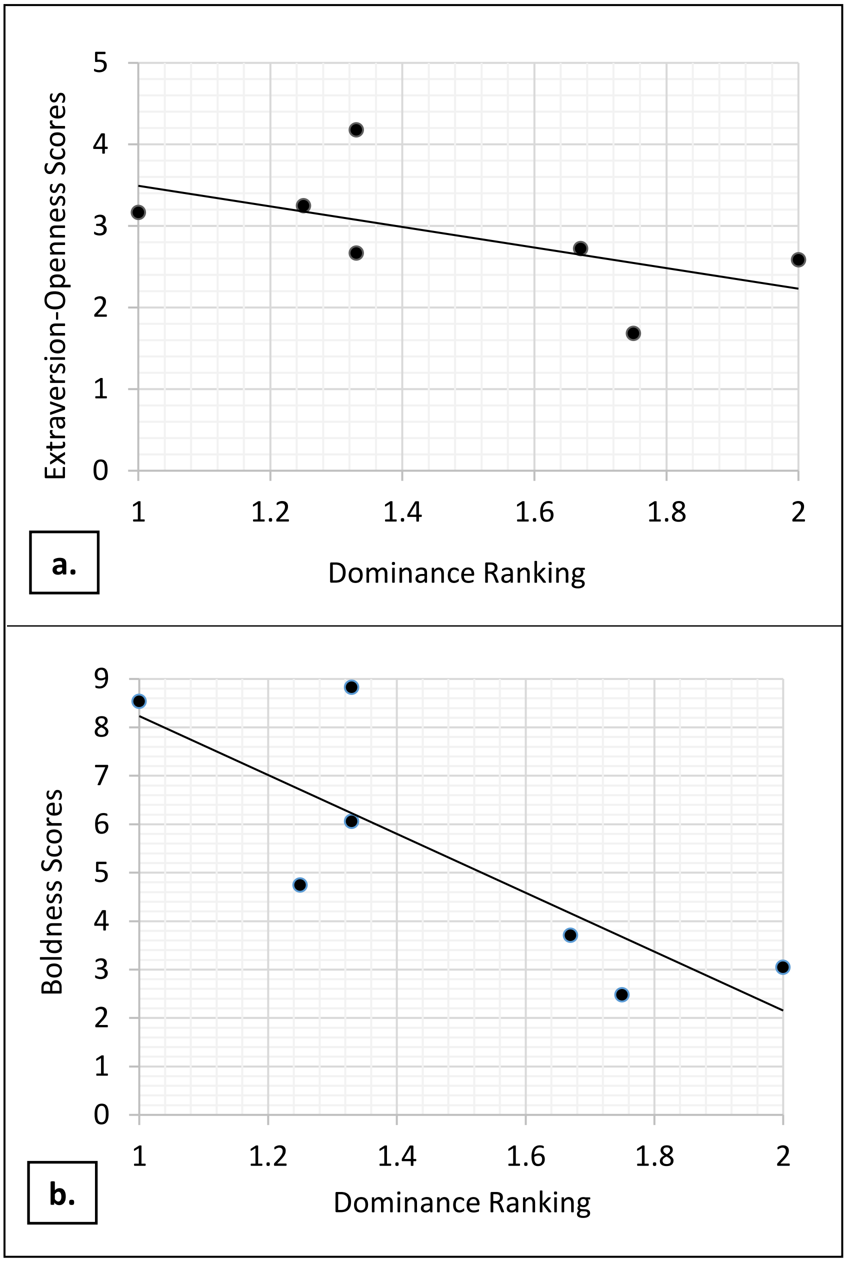

Figure A3. Harbor seal dominance rankings plotted against scores on a. Extraversion-Openness, and b. Boldness.

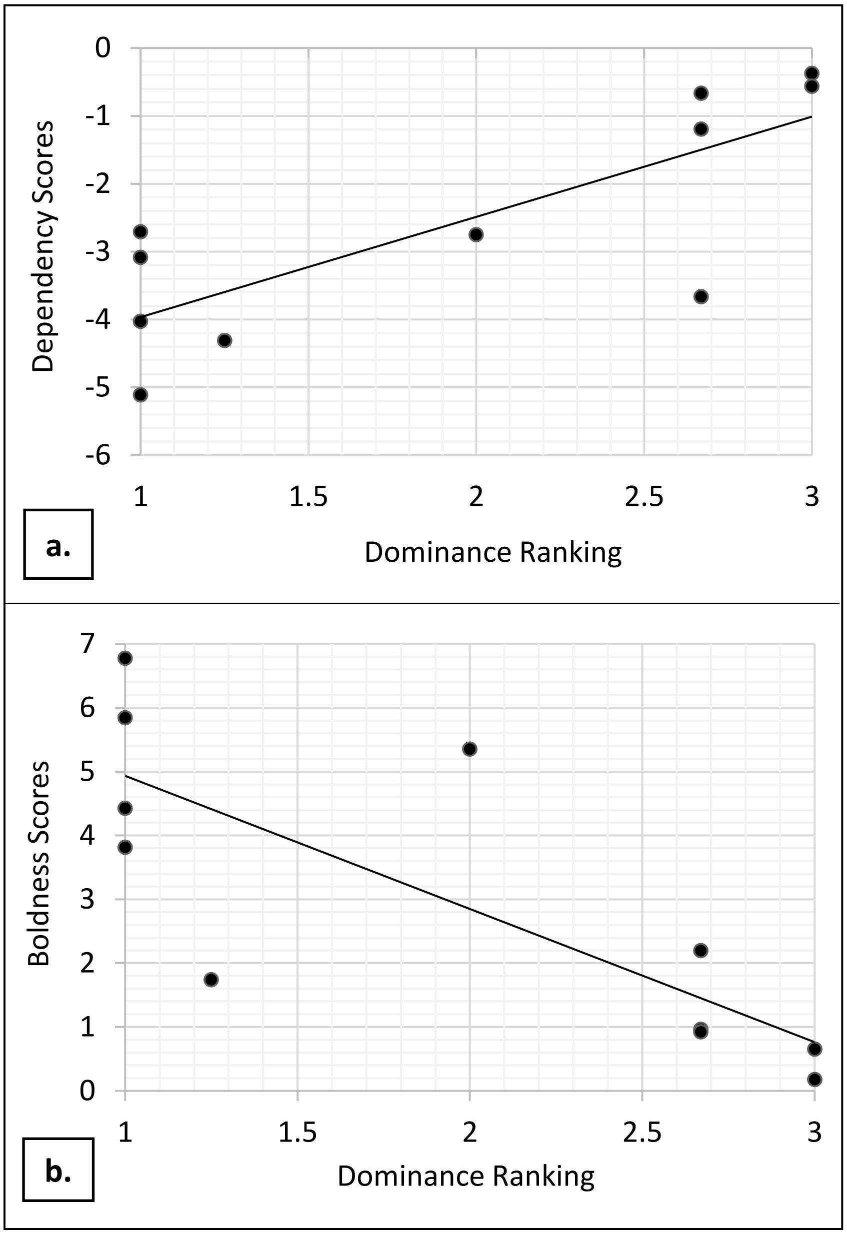

Figure A4. California sea lion dominance plotted against scores on a. Dependency, and b. Boldness.

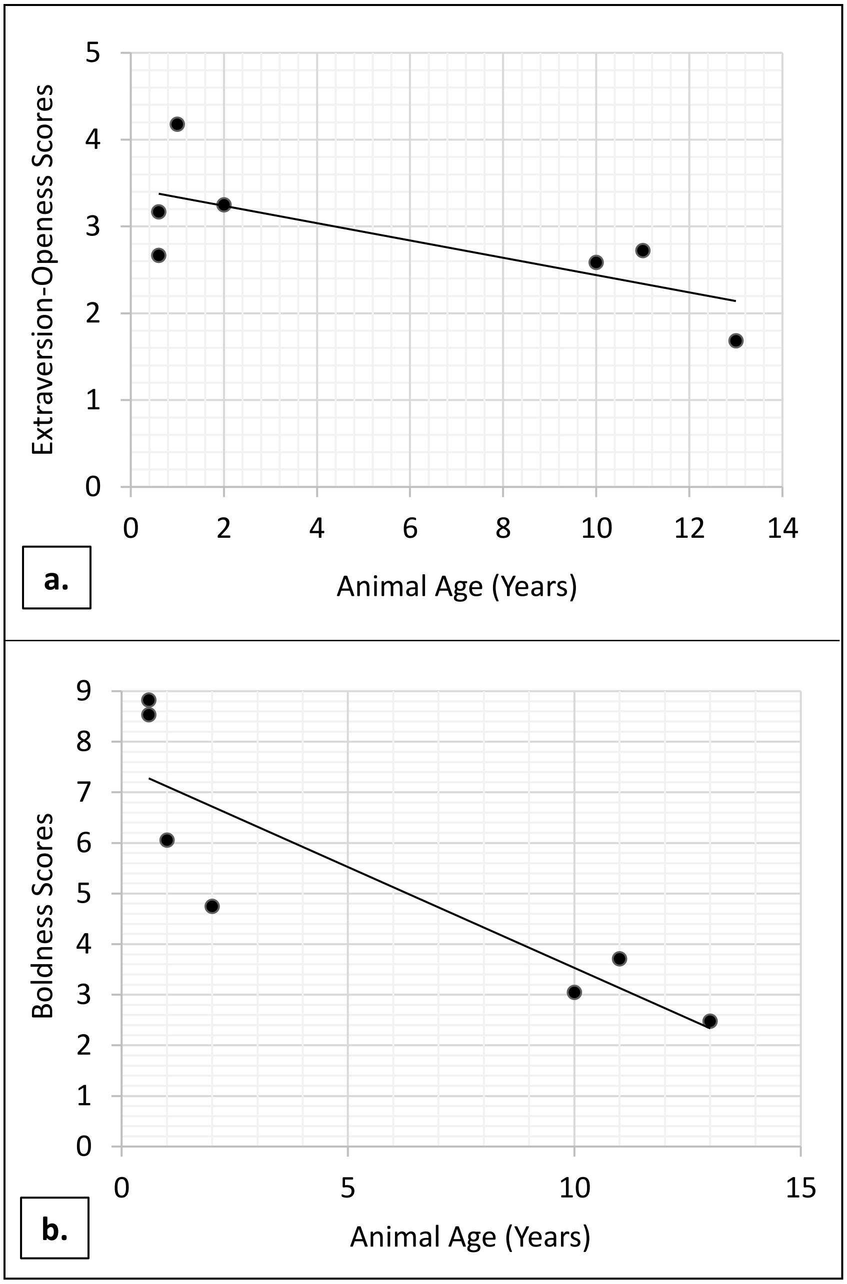

Figure A5. Harbor seal age plotted against scores on a. Extraversion-Openness, and b. Boldness.

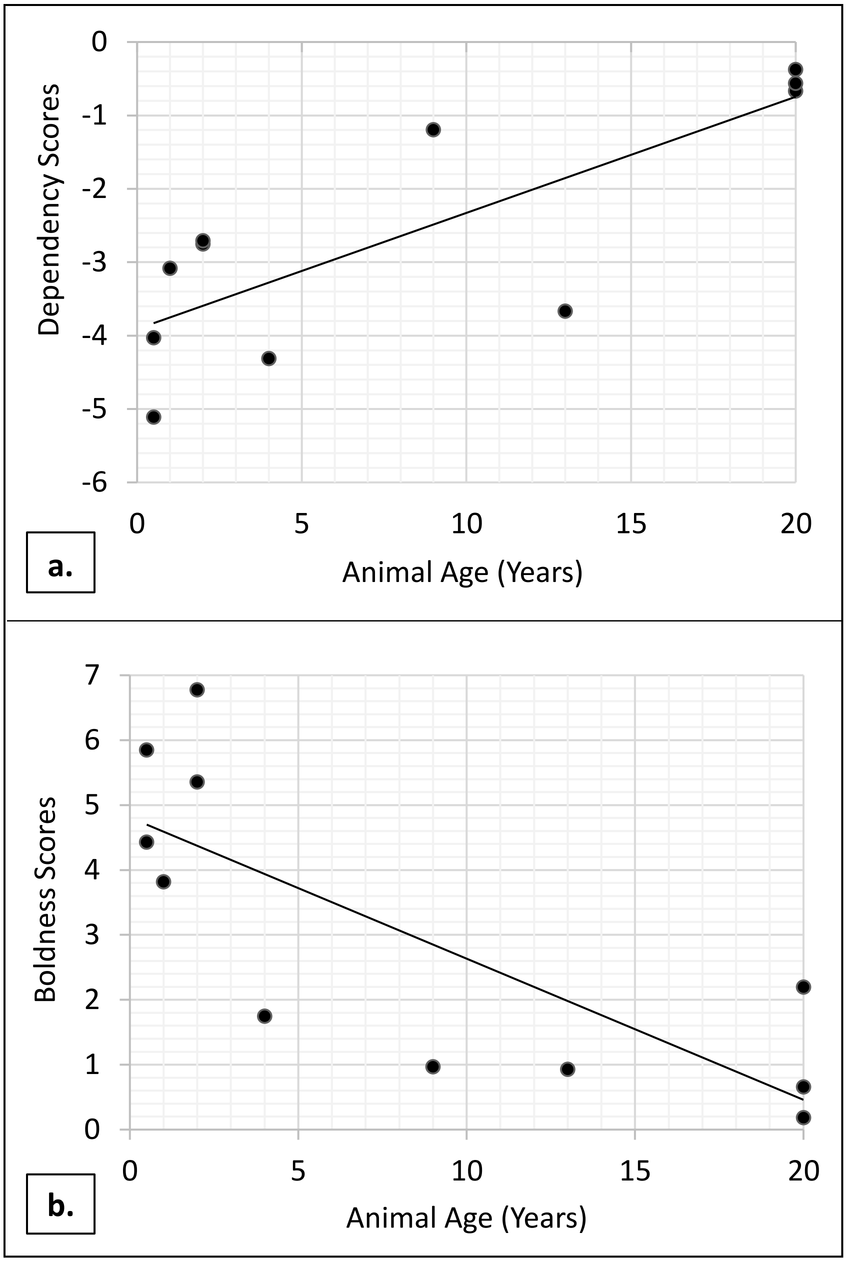

Figure A6. California sea lion age plotted against scores on a. Dependency, and b. Boldness.

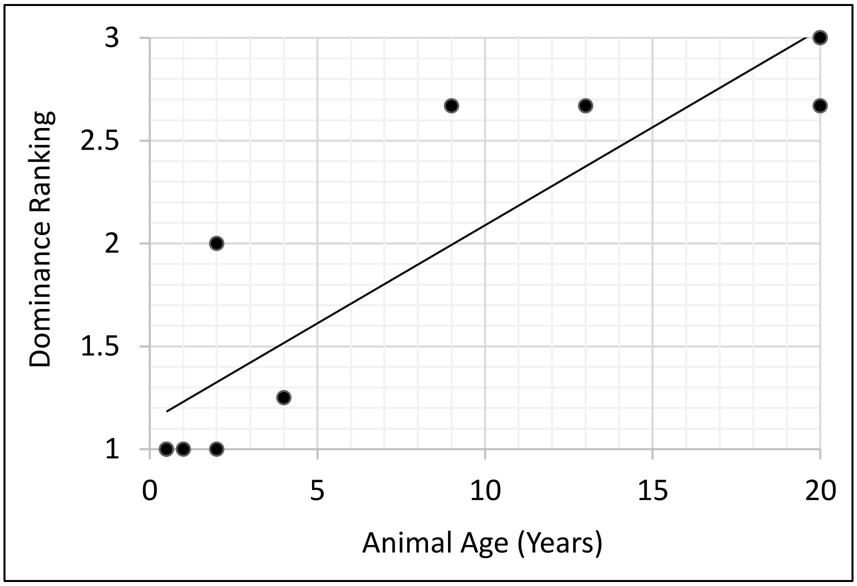

Figure A7. California sea lion dominance rankings plotted against animal age.

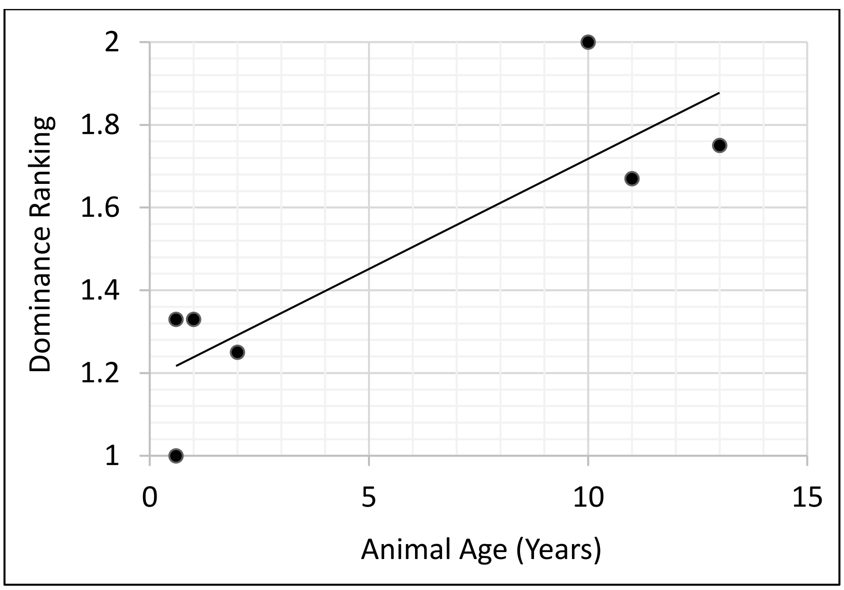

Figure A8. Harbor seal dominance rankings plotted against animal age.

APPENDIX B – IRB Approval Letter

THE UNIVERSITY OF SOUTHERN MISSISSIPPI

INSTITUTIONAL REVIEW BOARD
118 College Drive #5147 | Hattiesburg, MS 39406-0001
Phone: 601.266.5997 | Fax: 601.266.4377 | www.usm.edu/research/institutional.review.board

NOTICE OF COMMITTEE ACTION

The project has been reviewed by The University of Southern Mississippi Institutional Review Board in accordance with Federal Drug Administration regulations (21 CFR 26, 111), Department of Health and Human Services (45 CFR Part 46), and university guidelines to ensure adherence to the following criteria:

- The risks to subjects are minimized.
- The risks to subjects are reasonable in relation to the anticipated benefits.
- The selection of subjects is equitable.
- Informed consent is adequate and appropriately documented.
- Where appropriate, the research plan makes adequate provisions for monitoring the data collected to ensure the safety of the subjects.
- Where appropriate, there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of all data.
- Appropriate additional safeguards have been included to protect vulnerable subjects.
- Any unanticipated, serious, or continuing problems encountered regarding risks to subjects must be reported immediately, but not later than 10 days following the event. This should be reported to the IRB Office via the "Adverse Effect Report Form".
- If approved, the maximum period of approval is limited to twelve months. Projects that exceed this period must submit an application for renewal or continuation.

PROTOCOL NUMBER: 16031002
PROJECT TITLE: Assessing Personality in California Sea Lions and Harbor Seals using Trait Rating and Coding, Incorporating Ratings of Affective States
PROJECT TYPE: New Project
RESEARCHER(S): Amber de Vere
COLLEGE/DIVISION: College of Education and Psychology
DEPARTMENT: Psychology
FUNDING AGENCY/SPONSOR: N/A
IRB COMMITTEE ACTION: Expedited Review Approval
PERIOD OF APPROVAL: 03/11/2016 to 03/10/2017
Lawrence A. Hosman, Ph.D.
Institutional Review Board

REFERENCES

Arnason, U., Gullberg, A., Janke, A., Kullberg, M., Lehman, N., Petrov, E.A., Vainola, R. (2006). Pinniped phylogeny and a new hypothesis for their origin and dispersal. *Molecular Phylogenetics and Evolution*, 41(2), 345-354.

Barnard, A., Marshall-Pescini, S.M., Passalacqua, C., Beghelli, V., Capra, A., Normando, S., Pelosi, A., Valsecchi, P. (2016). Does subjective rating reflect behavioural coding? Personality in 2 month-old dog puppies: An open-field test and adjective-based questionnaire. *PLoS One*, 11(3), E0149831.

Bell, A.M., Hankison, S.J., Laskowski, K.L. (2009). The repeatability of behavior: a meta-analysis. *Animal Behavior*, 77(4), 771-783.

Broom, D.M. (1993). Assessing the welfare of modified or treated animals. *Livestock Production Science,* 36, 39-54.

Bornstein, M.H., Giusti, Z., Leach, D. B., Venuti, P. (2005). Maternal reports of adaptive behaviours in young children: Urban-rural and gender comparisons in Italy and the United States. *Infant and Child Development*, 14, 403-424.

Buck, R. (1999). The biology of affects: A typology. Psychological Review, 106, 301-336.

Campbell, D.T., Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105.

Cicchetti, D.V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284-290.

Ciardelli, L., Weiss, A., Powell, D., Reiss, D., Call, J, Fragaszy, D.M. (2017). Personality Dimensions of the Captive California Sea Lion (*Zalophus californianus*). *Journal of Comparative Psychology,* 131(1), 50-58.

Costa, P.T., McCrae, R.R. (1980). Influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. *Journal of Personality and Social Psychology*, 38(4), 668-678.

Dawkins, M.S. (2000). Animal minds and animal emotions. *American Zoology*, 40, 883-888.

de Vere, A.J., Levine, R. (under review). Human perception of the emotional repertoires of harbour seals (*Phoca vitulina*) and California sea lions (*Zalophus californianus*). *Animals & Society*.

de Vere, A.J., Lilley, M.K., Highfill, L. (2017). Do pinnipeds have personality? Broad dimensions and contextual consistency of behavior in harbor seals (*Phoca vitulina*) and California sea lions (*Zalophus californianus*). *International Journal of Comparative Psychology,* 30, 1-15.

de Vere, A.J., Kuczaj, S.A. (2016). Where are we in the study of animal emotions? *Wiley Interdisciplinary Review Cognitive Science*, 7(5), 354-362.

DeYoung, C.G., Quilty, L.C., Peterson, J.B. (2007). Between facets and domains: 10 aspects of the big five. *Journal of Personality and Social Psychology*, 93(5), 880-896.

Eckardt, W., Steklis, H.D., Steklis, N.G., Fletcher, A.W., Stoinski, T.S., Weiss, A. (2015). Personality Dimensions and Their Behavioral Correlates in Wild Virunga Mountain Gorillas (Gorilla beringei beringei). *Journal of Comparative Psychology,* 129(1), 26-41.

Feaver, J., Mendl, M., Bateson, P. (1986). A method for rating the individual distinctiveness of domestic cats. *Animal Behaviour*, 34, 1016-1025.

Freeman, H., Brosnan, S.F., Hopper, L.M., Lambeth, S.P., Schapiro, S.J., Gosling, S.D. (2013). Developing a comprehensive and comparative questionnaire for measuring personality in chimpanzees using a simultaneous top-down/bottom-up design. *American Journal of Primatology,* 75(10), 1042-1053.

Freeman, H., Gosling, S.D., Schapiro, S.J. (2011). Methods for assessing personality in non-human primates. In: A. Weiss, J. King, L. Murray (Eds.), *Personality and behavioral syndromes in nonhuman primates* (pp. 17-41). New York: Springer.

Frick, E. (2016). Establishing a link between personality and social rank in a group of bottlenose dolphins (Tursiops truncatus). *ProQuest Dissertations and Theses*.

Gartner, M.C., Powell, D.M., Weiss, A. (2016). Comparison of subjective well-being and personality assessments in the clouded leopard (*Neofelis nebulosa*), snow leopard (*Panthera uncia*), and African lion (*Panthera leo*). *Journal of Applied Animal Welfare Science*, 19(3), 294-302.

Godsell, J. (1988). Herd formation and haul-out behaviour in harbor seals (*Phoca vitulina*). *Journal of Zoology*, 215(1), 83-98.

Godsell, J. (1991). The relative influence of age and weight on the reproductive behaviour of male grey seals *Halichoerus grypus*. *Journal of Zoology*, 224, 537-551.

Goldberg, L. (1990). An Alternative "Description of Personality": The Big-Five Factor Structure. *Journal of Personality and Social Psychology*, 59(6), 1216-1229.

Gosling, S. (2001). From mice to men: what can we learn about personality from animal research? *Psychology Bulletin*, 127, 45–86.

Gosling, S., John, O. (1999). Personality Dimensions in Nonhuman Animals: A Cross-Species Review. *Current Directions in Psychological Science*, 8, 69-75.

Gottlieb, D., Capitanio, J., Mccowan, B. (2013). Risk factors for stereotypic behavior and self-biting in rhesus macaques (*Macaca mulatta*): Animal's history, current environment, and personality. *American Journal of Primatology*, 75(10), 995-1008.

Hayes, A.F., Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1, 77-89.

Heath, C.B., Perrin, W.F. (2009). C – California, Galapagos, and Japanese sea lions: *Zalophus californianus, Z. wollebaeki,* and *Z. japonicus*. In W.F. Perrin, B. Wursig, J.G.M. Thewissen (Eds.), Encyclopedia of Marine Mammals (Second Edition) (pp.170-176). Toronto, Canada: Elsevier Science.

Highfill, L., Hanbury, D., Kristiansen, R., Kuczaj, S., Watson, S. (2010). Rating vs. Coding in Animal Personality Research. *Zoo Biology*, 29, 509-516.

Highfill, L., Kuczaj, S. (2007). Do Bottlenose Dolphins (Tursiops truncates) Have Distinct and Stable Personalities? *Aquatic Mammals*, 33, 380-389.

Horback, K.M., Miller, L.J., Kuczaj, S.A. (2013). Personality assessment in African elephants (*Loxodonta africana*): Comparing the temporal stability of ethological coding versus trait rating. *Applied Animal Behaviour Science*, 149(1-4), 55-62.

Horn, J.L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179–185. http://dx.doi.org/10.1007/ BF02289447

Izard, C.E., Libero, D.Z., Putnam, P., Haynes, O.M. (1993). Stability of emotion experiences and their relations to traits of personality. *Journal of Personality and Social Psychology*, 64(5), 847-860.

John, O.P., Caspi, A., Robins, R.W., Moffitt, T.E., Stouthamer-Loeber, M. (1994). The “Little Five”: Exploring the Nomological Network of the Five-Factor Model of Personality in Adolescent Boys. *Child Development*, 65(1), 160-178.

Johnson-Laird, P.N., Oatley, K. (1989). The language of emotions: An analysis of a semantic field. *Cognition and Emotion*, 3, 81-123.

Jung, S., Lee, S. (2011). Exploratory factor analysis for small samples. *Behavior Research Methods*, 43, 701–709.

King, J., Figueredo, A. (1997). The Five-Factor Model plus Dominance in Chimpanzee Personality. *Journal of Research in Personality*, 31, 257-271.

King, J.E., Landau, V.I. (2003). Can chimpanzee (*Pan troglodytes*) happiness be estimated by human raters? *Journal of Research in Personality*, 37(1), 1–15.

King, J.E., Weiss, A., Farmer, K.H. (2005). A Chimpanzee (*Pan troglodytes*) Analogue of Cross-National Generalization of Personality Structure: Zoological Parks and an African Sanctuary. *Journal of Personality*, 73(2), 389-410.

King, J.E., Weiss, A., Sisco, M.M. (2008). Aping humans: Age and sex effects in chimpanzee (*Pan troglodytes*) and human (*Homo sapiens*) personality. *Journal of Comparative Psychology*, 122(4), 418-427.

Kuczaj, S.A., Highfill, L., Byerly, H. (2012). The importance of considering context in the assessment of personality characteristics: evidence from ratings of dolphin personality. *International Journal of Comparative Psychology*, 25, 309-329.

Kuczaj, S.A., Highfill, L.E., Makecha, R.N., Byerly, H.C. (2013). Why do dolphins smile? A comparative perspective on dolphin emotions and emotional expressions. In: Watanabe S, Kuczaj SA (eds) *Comparative perspectives on human and animal emotions*, Springer, Tokyo. pp 63-86.

Kuppens, P. (2005). Interpersonal determinants of trait anger: low agreeableness, perceived low social esteem, and the amplifying role of the importance attached to social relationships. *Personality and Individual Differences*, 38(1), 13-23.

Lilley, M.K. de Vere, A.J., Yeater, D.B, Kuczaj, S.A. (2018). Characterizing Curiosity-related Behavior in Bottlenose (*Tursiops truncatus*) and Rough-toothed (*Steno bredanensis*) Dolphins. *International Journal of Comparative Psychology*, 31, 1-22.

Lloyd, A.S., Martin, J.E., Bornett-Gauci, H.L.I., Wilkinson, R.G. (2007). Evaluation of a novel method of horse personality assessment: Rater-agreement and links to behaviour. *Applied Animal Behaviour Science*, 105(1-3), 205-222.

Lloyd, A.S., Martin, J.E., Bornett-Gauci, H.L.I., Wilkinson, R.G. (2008). Horse personality : Variation between breeds. *Applied Animal Behaviour Science*, 112(3-4), 369-383.

Martau, P.A., Caine, N.G., Candland, D.K. (1985). Reliability of the Emotions Profile Index, primate form, with *Papio hamadryas, Macaca fuscata,* and two *Saimiri* species. *Primates*, *26*, 501–505.

Marten, K., Psarakos, S. (1995). Evidence of self-awareness in the bottlenose dolphin (*Tursiops truncates*). In S. Taylor-Parker, R. W. Mitchell, & M. L. Boccia (Eds.), *Self-awareness in animals and humans: Developmental perspectives* (pp. 361-379). New York: Cambridge University Press.

McCrae, R.R. (2007). Aesthetic Chills as a Universal Marker of Openness to Experience. *Motivation and Emotion*, 31(1), 5-11.

Mehrabian, A. (1996). Analysis of the big-five personality factors in terms of the PAD temperament model. *Australian Journal of Psychology*, 48(2), 86-92.

Moreno, K.R., Highfill, L., Kuczaj, S.A. (2017). Does personality similarity in bottlenose dolphin pairs influence dyadic bond characteristics? *International Journal of Comparative Psychology*, 30, 1-15.

Morris, P.H., Doe, C., Godsell, E. (2008). Secondary emotions in non-primate species? Behavioural reports and subjective claims by animal owners. *Cognition and Emotion*, 22(1), 3-20.

Morris, P.H., Gale, A., Duffy, K. (2002). Can judges agree on the personality of horses? *Personality and Individual Differences*, 33, 67-81.

Morton, F. B., Lee, P. C., Buchanan-Smith, H. M., Brosnan, S. F., Thierry, B., Paukner, A.,...Weiss, A. (2013). Personality structure in brown capuchin monkeys (Sapajus apella): Comparisons with chimpanzees (Pan troglodytes), orangutans (Pongo spp.), and rhesus macaques (Macaca mulatta). *Journal of Comparative Psychology*, 127, 282–298.

Neumann, D. (1999). Agonistic behavior in harbor seals (*Phoca vitulina*) in relation to the availability of haul-out space. *Marine Mammal Science*, 15(2), 507-525.

Nunnally, J.C. (1978). Psychometric theory (2nd Ed.). New York: McGraw-Hill.

Ortony, A., Clore, G.L., & Foss, M.A. (1987). The referential structure of the affective lexicon. *Cognitive Science*, 11(3), 341-364.

Panksepp, J. (2011). The basic emotional circuits of mammalian brains: do animals have affective lives? *Neuroscience & Biobehavioral Reviews*, 35(9), 1791-1804.

Paul, E.S., Harding, E.J., Mendl, M. (2005). Measuring emotional processes in animals: the utility of a cognitive approach. *Neuroscience and Biobehavioural Reviews*, 29(3), 469-491.

Pederson, A.K., King, J.E., Landau, V.I. (2005). Chimpanzee (*Pan troglodytes*) personality predicts behavior. *Journal of Research in Psychology*, 39, 534-549.

Peterson, R.S., Bartholomew, G.A. (1967). The natural history and behavior of the California sea lion. *Special Publication, No. 1*. Stillwater, Oklahoma: The American Society of Mammologists.

Preston, S.D., de Waal, F.B.M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, 25, 1-72.

Pytlik Zillig, L.M., Hemenover, S.H., Dienstbier, R.A. (2002). What do we assess when we assess a big 5 trait?: A content analysis of the affective, behavioral, and cognitive processes represented in big 5 personality inventories. *Personality and Social Psychology Bulletin*, 28(6), 847-858.

Riedman, M. (1991). The Pinnipeds: Seals, Sea Lions, and Walruses. Oakland, California: University of California Press.

Roberts, B.W., Walton, K.E., Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin*, 132(1), 1-25.

Robinson, K., Twiss, S., Hazon, N., Moss, S., Lonergan, M., Pomeroy, P. (2015). Conspecific recognition and aggression reduction to familiars in newly weaned, socially plastic mammals. *Behavioral Ecology & Sociobiology*, 69, 1383-1394.

Russell, J.A. (2003). Core affect and the psychological construction of emotion. Psychological Review, 110, 145–172.

Schusterman, R.J. (1968). Experimental studies of pinniped behaviour. In R.J. Harrison, R.C. Hubbard, R.S. Peterson, C.E. Rice, R.J. Schusterman (Eds.), *The behaviour and physiology of pinnipeds* (pp. 87-171). New York: Appleton-Century-Crofts.

Stevens, J., Thyssen, A., Laevens, H., Vervaecke, H. (2013). The influence of zoo visitor numbers on the behavior of harbor seals (Phoca vitulina). *Journal of Zoo and Aquarium Research*, 1, 31-34.

Sullivan, R.M. (1982). Agonistic behavior and dominance relationships in the harbour seal, *Phoca vitulina*. *Journal of Mammalogy*, 63(4), 554-569.

Sussman, A., Ha, J., Bentson, K., Crockett, C. (2013). Temperament in Rhesus, Long-Tailed, and Pigtailed Macaques Varies by Species and Sex. *American Journal of Primatology,* 75(4), 303-313.

Tellegen, A. (1985). Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In: Hussain, T.A., Maser, J.D. (Eds.) *Anxiety and the anxiety disorders* (pp. 681-706). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc..

Twiss, S.D., Franklin, J. (2010) Individually Consistent Behavioral Patterns in Wild, Breeding Male Grey Seals (*Halichoerus grypus*). *Aquatic Mammals*, 36(3), 234-238.

Twiss, S.D., Culloch, R., Pomeroy, P.P. (2011). An in-field experimental test of pinniped behavioral types. *Marine Mammal Science*, 28(3), 280-294.

Twiss, S.D., Cairns, C., Culloch, R.M., Richards, S.A., Pomeroy, P.P. (2012). Variation in Female Grey Seal (Halichoerus grypus) Reproductive Performance Correlates to Proactive-Reactive Behavioral Types. *PLoS One*, 7(11), e49598, doi: 10.1371/journal.pone.0049598.

Uher, J., Asendorpf, J.B. (2008). Personality assessment in the Great Apes: Comparing ecologically valid behavior measures, behavior ratings, and adjective ratings. *Journal of Research in Personality*, 42(4), 821-838.

Uher, J., Werner, C.S., Gosselt, K. (2013). From observations of individual behaviour to social representations of personality: Developmental pathways, attribution biases, and limitations of questionnaire methods. *Journal of Research in Personality, 47*(5), 647-667.

Vazire, S., Carlson, E.N. (2010). Self-knowledge of personality: Do people know themselves? *Social and Personality Compass*, 4(8), 605-620.

Vazire, S., Gosling, S.D., Dickey, A.S., Schapiro, S.J. (2007). Measuring Personality in Nonhuman Animals, In: Robins, R., Fraley, R., Krueger, R. (Eds.) *Handbook of Research Methods in Personality Psychology*, Guildford Press: New York.

Von Borell, C., Kulik, L., Widdig, A. (2016). Growing into the self: the development of personality in rhesus macaques. *Animal Behaviour*, 122, 183-195.

Watson, D., Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin,* 98(2), 219-235.

Watson, D., Clark, L.A., Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063-1070.

Watson, D., Suls, J., Haig, J. (2002). Global self-esteem in relation to structural models of personality and affectivity. *Journal of Personality and Social Psychology*, 83(1), 185-197.

Weiss, A., et al. (2007). A cross- setting study of chimpanzee (Pan troglodytes) personality structure and development: zoological parks and Yerkes National Primate Research Center. *American Journal of Primatology*, 69, 1264–1277.

Weiss, A., King, J.E., Hopkins, W.D. (2009). Assessing chimpanzee personality and subjective well-being in Japan. *American Journal of Primatology*, 71(4), 283-292.

Weiss, A., King, J.E., Murray, L. (2011). Springer Extras for Personality and Temperament in Nonhuman Primates. New York: Springer. Available at http://extras.springer.com/2011/978-1-4614-0175-9. Accessed January 10th, 2018.

Wemelsfelder, F. (1999). The problem of animal subjectivity and its consequences for the scientific measurement of animal suffering. In F. Dolin (Ed.), *Attitudes to animals* (pp. 37-53). Cambridge, UK: Cambridge University Press

Wiener, P., Haskell, M.J. (2016). Use of questionnaire-based data to assess dog personality. *Journal of Veterinary Behavior: Clinical Applications and Research*, 16, 81-85.

Zevon, M.A., Tellegen, A. (1982). The structure of mood change: An idiographic/nomothetic analysis. *Journal of Personality and Social Psychology*, 43(1), 111-122.